Università degli Studi Sapienza di Roma
Dipartimento di Scienze Sociali ed Economiche
Dottorato di Ricerca in Statistica Economica
XXII Ciclo

Economic Growth and Technology Diffusion in a Continuous Time Model

Dottoranda:
Annarita Cardoni

Relatore:
Prof. Bernardo Maggi

A dissertation submitted to the Sapienza Università di Roma
in partial fulfillment of the requirements for the degree
of

Doctor of Philosophy in Economic Statistics

16 Dicembre 2014

ECONOMIC GROWTH AND TECHNOLOGY DIFFUSION IN A CONTINUOUS TIME MODEL

Contents

CHAPTER 1: Continuous time dynamic economic models

1.1 Introduction

This chapter deals with specification and econometric estimation of continuous time dynamic economic models. It relies on Gandolfo (1981) which provides an excellent overview of these aspects. This should help to better understand the continuous time model based on economic growth and technology diffusion shown in the next chapters.

Moreover, given the use of adjustment equations in the next chapters, in the next paragraph we shall recall some preliminary notions concerning these equations.

1.2 A digression on partial adjustment equations

In a *distributed lag equation* the value that a variable y takes at time t is function (for simplicity assumed linear) of the present and past values of some other variable x:

$$y_t = b_0 x_t + b_1 x_{t-1} + b_2 x_{t-2} + ... + b_n x_{t-n} + ..., \tag{1}$$

where the b's are known constants.

Distributed lag equations arise in many different situations. For example when changes in a variable x have their effect only gradually on y, or when macroeconomic variables react more or less slowly to changes in policy instruments (money supply, government expenditure, etc.). In all these cases, distributed lag equations constitute a sound representation of these relations.

The sum of the coefficients in equation (1) is assumed to be finite and its number of terms may be finite or infinite. Hence we have

$$b_0 + b_1 + b_2 + ... = b, \tag{2}$$

where b is a finite constant. Generally it is assumed that the b_i are non-negative.

If we define a new set of coefficients w_i

$$w_i = \frac{b_i}{b}, \tag{3}$$

then equation (1) can be rewritten as

$$y_t = b(w_0 x_t + w_1 x_{t-1} + w_2 x_{t-2} + ... + w_n x_{t-n} + ...), \tag{4}$$

where $\sum_{i=0}^{\infty} w_i = 1$. Observing equation (4) we can easily affirm that if each value of x had been increased by one unit, the value of y_t would increase by b units. This means that all values of x increase by one unit and maintain their new value through time (long term effect of a sustained unit increment in x). Thus b is called the *long term distributed lag multiplier*; the coefficient b_0 is called *impact multiplier* and the coefficients b_i (i=1,2,...) are called *delay - i multipliers.*

Let us now define the *mean time lag* as the weighted arithmetic mean of the time lags, where the weights are the coefficients w_i

$$0 * w_0 + 1 * w_1 + 2 * w_2 + ... + n * w_n + ... = \sum_{i=0}^{\infty} i w_i. \tag{5}$$

This is equivalent to write

$$\frac{\sum_{i=0}^{\infty} i b_i}{\sum_{i=0}^{\infty} b_i} = \frac{\sum_{i=0}^{\infty} i b_i}{b}. \tag{6}$$

Note that as the lags are amounts of time, their mean is an amount of time and this explains its denomination.

There exists a certain type of lag distribution equation equivalent to a widely used *partial adjustment equation.* According to it, given a discrepancy between the desired and the actual value of a variable, the latter is adjusted towards the former only gradually, according to a coefficient of reaction. This lag distribution equation will be largely used in the next chapters.

A particular lag distribution is the *geometric distributed lag equation* developed by Koyck, where the coefficients b_0, b_1, b_2,... decrease in a given ratio k in a infinite geometric series

$$b_i = kb_{i-1}, \qquad i = 1, 2, ..., \qquad 0 < k < 1. \tag{7}$$

Equation (7) can be considered as a first-order difference equation, the solution of which is

$$b_i = b_0 k^i. \tag{8}$$

The sum of the infinite geometric series of coefficients is

$$b = \sum_{i=0}^{\infty} b_i = \sum_{i=0}^{\infty} b_0 k^i = b_0 \sum_{i=0}^{\infty} k^i = b_0 \frac{1}{1-k};^{1} \tag{9}$$

substituting in (8) and considering that from (9) $b_0 = b(1-k)$, we obtain

$$b_i = b(1-k)k^i, \tag{10}$$

[1]Note that $\sum_{i=0}^{\infty} k^i$ is a geometric series converging to $\frac{1}{1-k}$ as $0 < k < 1$.

and so the first coefficient, b_0, is fixed as *b(1-k)* in order that the sum of the infinite geometric series of coefficients is b. Therefore equation (1) can be written as

$$y_t = b(1-k)(x_t + kx_{t-1} + k^2 x_{t-2} + ...). \tag{11}$$

Koych scheme is very useful because it can be reduced to a relationship involving y_t, x_t and y_{t-1} only. In fact, shifting all the time subscripts backwards by one unit and multiplying throughout by k we can write

$$ky_{t-1} = b(1-k)(kx_{t-1} + k^2 x_{t-2} + ...). \tag{12}$$

Subtracting (12) from (11) and eliminating the common terms, we have

$$y_t - ky_{t-1} = b(1-k)x_t, \tag{13}$$

therefore

$$y_t = b(1-k)x_t + ky_{t-1}. \tag{14}$$

This equation can be manipulated still further, in fact, by subtracting y_{t-1} from both members we obtain

$$\begin{aligned} y_t - y_{t-1} &= b(1-k)x_t + ky_{t-1} - y_{t-1} \\ &= b(1-k)x_t - y_{t-1}(-k+1) \\ &= (1-k)(bx_t - y_{t-1}) \\ &= (1-k)(bx_t - y_{t-1}), \end{aligned} \tag{15}$$

and if we further assume that the weights were normalized from the beginning in such a way that $b = \sum_{i=0}^{\infty} b_i = 1$, we finally have

$$y_t - y_{t-1} = (1-k)(x_t - y_{t-1}), \tag{16}$$

which is the typical form of the *partial adjustment equation*, where x_t is the "desired" or "potential" value of y_t. Hence, the generic partial adjustment equation

$$y_t - y_{t-1} = \alpha(\hat{y}_t - y_{t-1}), \qquad 0 < \alpha < 1, \tag{17}$$

is equivalent to a Koych distributed lag equation.

The results obtained in discrete time can be extended to continuous time.

A *continuously distributed lag equation* is

$$\begin{aligned} y_t &= \int_0^{\infty} [f(\tau)x(t-\tau)]d\tau, \qquad \int_0^{\infty} f(\tau)d\tau = 1, \\ \lim_{\tau \to \infty} f(\tau) &= 0, \end{aligned} \tag{18}$$

where $f(\tau)$ is the time-form of the weighting function; for simplicity we have assumed that the integral of the weights is one instead of an arbitrary positive constant. In the contrary case we would write $y_t = b\int_0^\infty [f(\tau)x(t-\tau)]d\tau$, $\int_0^\infty f(\tau)d\tau = \frac{1}{b}$.

The continuous counterpart of the geometric lag distribution is the *exponential lag distribution*, where

$$y_t = \int_0^\infty \alpha e^{-\alpha\tau} x(t-\tau)d\tau. \tag{19}$$

It is easy to check that

$$\int_0^\infty \alpha e^{-\alpha\tau} d\tau = 1. \tag{20}$$

Equation (19) is equivalent to the *partial adjustment equation in continuous time*

$$Dy_t = \alpha[x(t) - y(t)], \tag{21}$$

where $x(t)$ can be interpreted as the "desired" or "potential" value of y at time t, say $\hat{y}(t)$, and α is the speed of adjustment.

In fact, let us perform a change of variable from τ to s in the integral (19), defining $s = t - \tau$ (whence $\tau = t - s$). We have[2]

$$y_t = -\int_t^{-\infty} \alpha e^{-\alpha(t-s)} x(s)ds = \int_{-\infty}^t \alpha e^{-\alpha(t-s)} x(s)ds. \tag{22}$$

[2] According to the elementary rules of integral calculus and in particular to the formula for integration by substitution and the property of definite integrals, $\int_a^b \phi(\tau)d\tau = \int_c^d \phi[f(s)]f'(s)ds$, where $\tau = f(s)$, $a = \phi(c)$ and $b = \phi(d)$. In our case, $s = t$ when $\tau = 0$ and $s = -\infty$ when $\tau = \infty$. Further remember that the property according to which the definite integral changes sign when the limits of integration are reversed also holds when one of the limits is infinite provided that the integral converges.

As t is a constant with respect to the integral, $\alpha e^{-\alpha t}$ can be considered as a multiplicative constant; hence

$$y_t = \alpha e^{-\alpha t} \int_{-\infty}^{t} e^{\alpha s} x(s) ds, \tag{23}$$

whence

$$y_t e^{\alpha t} = \alpha \int_{-\infty}^{t} e^{\alpha s} x(s) ds. \tag{24}$$

Performing a differentiation with respect to time, we have

$$\alpha y(t) e^{\alpha t} + (Dy(t)) e^{\alpha t} = \alpha \frac{d}{dt} [\int_{-\infty}^{t} e^{\alpha s} x(s) ds]. \tag{25}$$

Differentiating the integral with respect to a parameter which occurs in the upper limit,[3] we can write

$$\frac{d}{dt} [\int_{-\infty}^{t} e^{\alpha s} x(s) ds] = e^{\alpha t} x(t), \tag{26}$$

therefore

$$\alpha y(t) e^{\alpha t} + e^{\alpha t} Dy(t) = \alpha e^{\alpha t} x(t), \tag{27}$$

[3]The derivative of the integral is equal to the derivative of the upper limit with respect to the parameter, multiplied by the integrand, in which the parameter has been substituted in the place of the variable of integration - remember the formula for the differentiation of a integral with respect to a parameter which occurs in both limits as well as in the integrand.

whence, eliminating $e^{\alpha t}$ and rearranging terms, we have

$$Dy(t) = \alpha[x(t) - y(t)], \tag{28}$$

which coincides with equation (21).

Let us continue to see what seen in discrete time also in continuous time.

In continuous time the *mean lag* is defined as

$$\int_0^\infty \tau f(\tau) d(\tau), \tag{29}$$

that is, with an exponential weighting function

$$\alpha \int_0^\infty \tau e^{-\alpha\tau} d(\tau). \tag{30}$$

It can be checked easily, taking into account the l'Hopital's rule, that

$$\int_0^\infty \tau e^{-\alpha\tau} d(\tau) = \frac{1}{\alpha^2}, \tag{31}$$

and substituting in (30), we have

$$\alpha \frac{1}{\alpha^2} = \frac{1}{\alpha}. \tag{32}$$

Hence when $\alpha \to +\infty$ the mean lag tends to 0, i.e. $y(t)$ tends to adjust immediately to $x(t)$.

The mean lag $\frac{1}{\alpha}$ can also be interpreted as the time necessary for about 63% of the discrepancy between $y(t)$ and $\hat{y}(t)$ to be eliminated by changes in $y(t)$ following a change in $\hat{y}(t)$. Consider the expression for $y(t)$ given by equation (22) and assume that θ units of time ago a change in x (that is in $\hat{y}$), say Δx, took place. Then the change in y is

$$\Delta y(t) = \int_{t-\theta}^{t} \alpha e^{-\alpha(t-s)} \Delta x d(s), \tag{33}$$

that is, as Δx is a given constant,

$$\Delta y(t) = \Delta x \int_{t-\theta}^{t} \alpha e^{-\alpha(t-s)} d(s). \tag{34}$$

Below, we can prove equation (33). Define a variable X which is equal to x for times farther away than θ units ago, and equal to $x + \Delta x$ for all times from $t - \theta$ up to t, that is

$$X = \begin{cases} x & for \quad s \leq t - \theta \\ x + \Delta x & for \quad t - \theta \leq s \leq t, \end{cases}$$

and let Y be the new value of y following the change in x. Then[4]

$$\begin{aligned} Y(t) &= \int_{-\infty}^{t} \alpha e^{-\alpha(t-s)} x(s) ds \\ &= \int_{-\infty}^{t-\theta} \alpha e^{-\alpha(t-s)} x(s) ds + \int_{t-\theta}^{t} \alpha e^{-\alpha(t-s)} (x(s) + \Delta x) ds, \end{aligned} \tag{35}$$

[4]The function $X(s)$ is continuous in the interval $-\infty$ to t except for a finite discontinuity at $s = t - \theta$. It is well known that if a function has a finite number of discontinuities in the interval of integration, then it is integrable over that interval.

and so

$$\begin{aligned}
\Delta y(t) &= Y(t) - y(t) \\
&= Y(t) - \int_{-\infty}^{t} \alpha e^{-\alpha(t-s)} x(s) ds \\
&= Y(t) - [\int_{-\infty}^{t-\theta} \alpha e^{-\alpha(t-s)} x(s) ds + \int_{t-\theta}^{t} \alpha e^{-\alpha(t-s)} x(s) ds]. \qquad (36)
\end{aligned}$$

Substituting $Y(t)$ from (35) into (36) and simplifying we have equation (33). In fact, we can write

$$\begin{aligned}
\Delta y(t) &= Y(t) - y(t) \\
&= \int_{-\infty}^{t-\theta} \alpha e^{-\alpha(t-s)} x(s) ds + \int_{t-\theta}^{t} \alpha e^{-\alpha(t-s)} [x(s) + \Delta x] ds + \\
& - \int_{-\infty}^{t-\theta} \alpha e^{-\alpha(t-s)} x(s) ds - \int_{t-\theta}^{t} \alpha e^{-\alpha(t-s)} x(s) ds \\
&= \int_{t-\theta}^{t} \alpha e^{-\alpha(t-s)} x(s) ds + \int_{t-\theta}^{t} \alpha e^{-\alpha(t-s)} \Delta x ds + \\
& - \int_{t-\theta}^{t} \alpha e^{-\alpha(t-s)} x(s) ds. \qquad (37)
\end{aligned}$$

Consider now equation (34), and go back to the original variable τ,[5] obtaining

$$\begin{aligned}
\Delta y &= \Delta x \int_{0}^{\theta} \alpha e^{-\alpha\tau} d\tau \\
&= \Delta x(-e^{-\alpha\theta} - (-1)) \\
&= \Delta x(1 - e^{-\alpha\theta}), \qquad (38)
\end{aligned}$$

[5]The procedure is a change of variable from s to $\tau = t - s$.

from which, letting the so far undetermined θ take on the value $\frac{1}{\alpha}$ and computing, we obtain $\Delta y \simeq 0.632\Delta x$.

Therefore $\theta = \frac{1}{\alpha}$ is the time necessary for about 63% of the discrepancy between $y(t)$ and $x(t)$ to be eliminated by changes in $y(t)$ following a change in $x(t)$. Of course, this interpretation holds when $x(t) = \hat{y}(t)$. This interpretation has to be kept in mind as it will be useful for the comprehension of the model shown in the next chapters.

1.3 Specification of the models

In this paragraph the main characteristics concerning the construction of dynamic models for the treatment of continuous economic phenomena will be shown. These models can be either single- or multi-equation models.

First of all the equations are usually specified as dynamic equations. This means that the model is recursive, in the sense that it is expressed as a system of differential equations in which the derivative of each endogenous variable is in principle a function of the levels of all the variables.

This has some advantages from the economic point. In fact, in this way the parameters of such models are identified and all the restrictions suggested by economic theory can be imposed directly on the model. Moreover, the effects of structural changes can be interpreted in terms of these parameters.

We will return on the econometric questions later and more deeply in paragraph 1.4 of chapter 1. Now, let us examine the advantages of using dynamic equations such as equation (21) of paragraph 1.2 of chapter 1 from the economic point of view.

To start with we observe that equation (21) includes both the case in which $x(t)$ is the "desired" or "potential" value of the variable which adjusts ($x(t) = \hat{y}(t)$: see paragraph 1.2 of chapter 1) - in which case we may speak of partial adjustment equations in the strict sense - and the case in which $x(t)$ is the desired variable of some other variable. More generally, we may suppose that the variable adjusts in relation to the discrepancy between the desired and actual values of other variables.

Thus we can write in general

$$Dy(t) = f[x(t) - v(t)], \tag{39}$$

where f is a sign-preserving function, and $f(0) = 0$, $f'(0) > 0$.

The variables $x(t)$ and $v(t)$ are to be defined; in the case in which $x(t) = \hat{y}(t)$ and $v(t) = y(t)$, as seen, we have a partial adjustment equation in the strict sense.

The logic underlying equations of type (39) can be based both on market mechanisms and on reaction functions of the policy makers. The market mechanisms may reflect adjustments to excess demand or the existence of obstacles, frictions and delays of various kinds that prevent the operators from responding to change instantaneously, so that they cannot move onto their schedule immediately (this is the case of partial adjustment proper). Equation (39) can be linearized, for example, in the form

$$Dy(t) = \alpha[x(t) - v(t)], \tag{40}$$

where the multiplicative constant α is $f'(0)$ and can be interpreted as a speed of adjustment or reaction coefficient. Only when the equation is a partial adjustment equation in the strict sense, α can be interpreted as examined in paragraph 1.2 of chapter 1.

Let us note that it would also be possible to make the adjustment coefficient a function of other variables. This means that the original specification should not be (39) but a function in which these other variables appear as arguments. As such a function should have the property of being zero when $x(t) = v(t)$ irrespective of the values of the other variables, the resulting form is the multiplicative one

$$Dy(t) = h(...)[x(t) - v(t)], \tag{41}$$

where $h(...)$ is a function of variables to be specified. Even under the usual simplifying assumption that h is a linear function, this adjustment equation would maintain a non-linear form.

In any case, the formulation of the model as a set of adjustment equations means admitting the possibility that the system is in a situation of disequilibrium. More specifically, we are in a context of disequilibrium dynamics, which seems more appropriate to real life phenomena. Of course, if on the basis of a priori information there is a strong presumption that a certain market is practically always in equilibrium, the equation relative to it can be written as an equilibrium, instead of an adjustment equation. In general, however, unless the a priori information is very sound, the formulation as an adjustment equation is preferable; if this is the case, econometric estimation will show, through a very high value of the adjustment coefficient, that the market under consideration is practically always in equilibrium.

Another important characteristic concerning the specification of continuous time dynamic economic models is that the model in principle ought to have an equilibrium, usually a steady state equilibrium.

It is important to note that this does not mean that the model is in this kind of equilibrium or that the equilibrium is necessarily stable. In other words, the adjustment processes are not a priori assumed to be such as to bring the model to its equilibrium. This seems an advantage respect to the usual assumption that the model must be stable. In fact, whether the model is stable or not can be ascertained with the econometric analysis, and, if the model is not stable and if the stability is believed an essential requisite, the cause of this instability can be searched modifying the model where appropriate.

The characteristic under consideration is simply that an equilibrium should exist. This may appear an obvious requisite, because the concept of equilibrium, and therefore its existence, is central to the economist's traditional way of thinking - apart from the question of its stability. But a model could be built leaving aside the concept of equilibrium, that is, by simply specifying dynamic behaviour equations and then solving the system to obtain the actual path or paths of the model, giving no privileged position to a particular equilibrium point or path.

Once the notion of equilibrium is accepted, it seem logical that in a dynamic model with the aim to describe the economy for a sufficiently long period of time, the equilibrium to be considered is a steady state equilibrium. In addition to this economic justification, there are other formal justifications. First of all, since the functions are usually non-linear, if the model has to be linearized, the equilibrium will be a convenient point where to perform such a linearization. Furthermore, if the analysis of the steady state is performed, its properties not only gives information on the dynamic behaviour of the model, but helps verifying the mathematical consistency of the model itself. In fact, implausible long-run behaviour could indicate a structural defect, such as the omission of an important feedback. If a model does not have a steady state, the variables will be fluctuating in some way for all t with unstable or explosive oscillations. It must however be pointed out that the presence of a steady state is not an indispensable characteristic, hence it is possible to build and estimate models which lack it or it is possible not having the necessity to study it.

Another purely formal characteristic relative to the specification of these models is that the adjustment equations are usually, but not always, formulated expressing the variables in terms of logarithms.

The last characteristic that it is worth to noticing is that the models considered are small in size and amenable not only to a quantitative analysis (numerical simulations, etc.) but also to a qualitative analysis (existence of the steady state, comparative dynamics, stability, sensitivity).

A limited number of equations helps skilling the difficulties due to the numerical problems or the mathematical difficulty of a complete qualitative analysis if performed. This last difficulty can increase as the number of equations increases, up to the point of becoming intractable. Therefore, the benefits of manageability and of analytical tractability have to be weighted against the benefits of a more detailed description. Nevertheless, small in size models can cover a broad range of economic relations and embody a sufficiently rich theoretical structure.

In conclusion, the model can be written as a first-order differential system in normal form

$$Dlog\boldsymbol{Y}(t) = \boldsymbol{H}[log\boldsymbol{Y}(t), \boldsymbol{Z}(t), \boldsymbol{\theta}] + \boldsymbol{u}(t), \tag{42}$$

where $\boldsymbol{Y}(t)$ is the vector of the endogenous variables, $\boldsymbol{Z}(t)$ the vector of the exogenous variable, $\boldsymbol{\theta}$ the vector of parameters which consists of a subvector $\boldsymbol{\alpha}$ of adjustments coefficients and a subvector $\boldsymbol{\beta}$ of other behavioural parameters (propensions, elasticities, etc.), $\boldsymbol{H}$ the vector of linear or non-linear differentiable functions, and $\boldsymbol{u}(t)$ the vector of disturbances with classic properties (white noise). The symbol D denotes the differential operator respect to time $\frac{d}{dt}$.

Let us finally note that the differential system could be in non-normal form or of an order higher than the first. In these cases, under certain conditions and by means of suitable manipulations, it can be reduced to a first order system in normal form.

1.4 Econometric estimation of the models

The models under consideration do not rise problems of identification,[6] as we saw in paragraph 1.3 of chapter 1. On the contrary, they are usually heavily overidentified, with the same parameters or functions of parameters appearing in several equation, and with cross-equation restrictions on the parameters. For this reason, estimation with full information methods can take adequate account of all these restrictions and hence allow to obtain efficient estimates despite the smallness of the samples. Among full information methods of estimation, full-information maximum likelihood (FIML) is preferable to the other major full-information technique (for example three-stage least squares), because it has the advantage of using a wide range of a priori information in the estimation procedure, pertaining not only to each equation individually but also to several equations simultaneously, such as non-linear within and across equation restrictions on the parameters. It is important to stress that, because of the sensitivity of full information methods to specification errors, the model will have to be specified with great care. Nevertheless, this sensitivity may be an advantage, because it helps to detect these errors, which might remain undetected when using other estimation methods.

[6]In our case of a system of differential equations the condition is that the system cannot be reduced to an equivalent system where any equation is of a lower order, so that the model is invariant under a linear transformation of time and the parameters of the differential equations may be identified.

Regarding the estimation procedure proper, that is the estimation of the original non-linear differential system directly, the approach consists in integrating the original non-linear model over the observation interval, in performing a non-linear discrete approximation to these integrals, and in estimating the resulting system of equations, which is a non-linear in variables and parameters simultaneous equation model. This procedure is very expensive, especially if the original differential system contains cross-equation restrictions on the parameters, so that a non-linear full information estimator is required. The question then arises as to whether better estimates of the true parameters are obtained using an exact discrete analogue to the linearized form of the original non-linear differential model or using an approximate non-linear discrete analogue to the same model.

In what follows we will see an estimation procedure involving a derivation of an exact and then of an approximate discrete analogue.

There would be problems and costs in the direct estimation of non-linear differential systems such that the use of linearized models could be required for the estimation of the parameters of these systems.[7] Concerning where to perform the linearization procedure, if the model has a steady state and if the sample period is sufficiently long for substantial changes in the variables to have occurred, linearization about the steady state may be more appropriate. It should be noted, however, that when the steady state is complicated and when the model is subject to modifications, this linearization is very time consuming because the steady state (and then the linearization) must be computed again after each modification. On the other hand, when the sample period is not too long and the changes in the variables are not substantial, or when the steady state is very complicated, it may be convenient to linearize about sample means. Nevertheless, even if the linearization about the steady state seems more appropriate and hence performed, the linearization about sample means may be used for a preliminary screening of the model: once a satisfactory specification has been determined, the linearization about the steady state will be used to obtain the final estimates. It should be pointed out that whichever linearization is employed, the pa-

[7] It is important to stress that the properties of FIML estimators of linear models are more developed than those for non-linear models but a non-linear estimator eliminates any bias arising from linearization and provides an estimate of any biases. Moreover, linearization may sometimes lead to parameters becoming unidentified, or poorly identified in that the asymptotic standard errors become very large; this is less likely with a non-linear estimator.

rameters being estimates are always those of the original non-linear model. Obviously, no problem of choice arises when the model has been constructed in such a way as not to possess a steady state.

In what follows we will cope with some questions considering a linearized system. Whichever linearization is adopted, we have to deal with a system of the type

$$D\boldsymbol{x}(t) = \boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{B}\boldsymbol{z}(t) + \boldsymbol{u}(t), \tag{43}$$

where the transformations to pass from $\boldsymbol{Y}$ to $\boldsymbol{x}$ and from $\boldsymbol{Z}$ to $\boldsymbol{z}$ depend on the non-linearities of the original system (see equation (42), in paragraph 1.3 of chapter 1) and on the point about which the linearization is performed. Of course the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ will be different, depending on the kind of linearization performed (about the steady state or about sample means). In the case of linearization about the steady state, the elements of these matrices will usually be complicate functions of the parameters, of the exogenous rates of growth and perhaps of the initial values of the exogenous variables. In the case of linearization about sample means, the elements of these matrices will usually be simple functions of the parameters and of the means of the variables. This is an undoubted practical advantage of the latter linearization over the former. Vector $\boldsymbol{u}(t)$ of equation (43) is a vector of disturbances assumed of white noise type. It is conventionally considered to satisfy the equation[8]

$$\boldsymbol{u}(t) = \frac{d}{dt}\boldsymbol{\zeta}(t), \tag{44}$$

where $\boldsymbol{\zeta}(t)$ is generated by a homogeneous stochastic process with uncorrelated increments and with mean zero, that is

[8]Assumption (44) can be dropped without consequences. Suppose that $\boldsymbol{u}(t)$ is generated by a stochastic process such that the integral $\boldsymbol{\zeta}(t) = \int_0^t \boldsymbol{u}(s)ds$ has the properties (45). Even if $\lim_{h\to 0} E[|\frac{\boldsymbol{\zeta}(t+h)-\boldsymbol{\zeta}(t)}{h} - \dot{\boldsymbol{\zeta}}(t)|^2]$ does not exist (so that $\boldsymbol{\zeta}(t)$ is non-m.s. -in mean square- differentiable and the process $\boldsymbol{u}(t)$ cannot be rigorously defined), it is possible to consider the system $d\boldsymbol{x}(t) = \boldsymbol{A}\boldsymbol{x}(t)dt + \boldsymbol{B}\boldsymbol{z}(t)dt + d\boldsymbol{\zeta}(t)$, where $\boldsymbol{u}(t)$ is replaced by the mean square differential of the process $\boldsymbol{\zeta}(t)$; the system above has the same properties as (43).

$$
\begin{aligned}
E[\boldsymbol{\zeta}(t)] &= 0, \quad \forall t, \\
E[\boldsymbol{\zeta}(t_1)-\boldsymbol{\zeta}(t_2)][\boldsymbol{\zeta}(t_3)-\boldsymbol{\zeta}(t_4)]^T &= 0, \quad \forall t_1 > t_2 \geq t_3 > t_4, \\
E[\boldsymbol{\zeta}(t+h)-\boldsymbol{\zeta}(t)][\boldsymbol{\zeta}(t+h)-\boldsymbol{\zeta}(t)]^T &= \boldsymbol{\Omega}|h|,
\end{aligned} \tag{45}
$$

where the second equation indicates the uncorrelated increments or orthogonality and the third equation indicates the homogeneity, and where $\boldsymbol{\Omega}$ is a matrix of constants.

We turn now to the estimation of model (43). We shall first consider the exact discrete model equivalent to (43) and then an approximate discrete analogue.

The basic idea is to represent the differential equation system by a stochastically equivalent system of difference equations such that this exact[9] discrete analogue is satisfied by any set of equispaced observations generated by the differential system. In this way, it will be possible to estimate the parameters of the model from a sample of discrete observations. The first step is to solve the differential system (43). As it is linear and non-homogeneous, its general solution is given by the general solution of the corresponding homogeneous[10] system plus a particular solution of the non-homogeneous system itself. Formally, the general solution of a system of the type

$$\boldsymbol{Dy}(t)=\boldsymbol{Ay}(t) + \boldsymbol{w}(t), \quad \boldsymbol{y}(0)=\boldsymbol{c}, \tag{46}$$

where $\boldsymbol{w}(t)$ is a vector of known functions, is given by

[9]"Exact" in the sense that it is stochastically equivalent, and not an approximation, to the differential system (43).

[10]"Homogeneous" and "non-homogeneous" are used here with the meaning that they have in the theory of differential equations. They should not be confused with the definitions concerning stochastic processes.

$$\boldsymbol{y}\textit{(t)} = \boldsymbol{r}\textit{(t)} + \int_0^t \boldsymbol{R}(t-\theta)\boldsymbol{w}(\theta)d\theta, \tag{47}$$

where $\boldsymbol{r}\textit{(t)}$ is the solution of $\boldsymbol{D}\boldsymbol{r}\textit{(t)}\textit{=}\boldsymbol{A}\boldsymbol{r}\textit{(t)}$, $\boldsymbol{r}\textit{(0)=}\boldsymbol{c}$, and $\boldsymbol{R}\textit{(t)}$ is the solution of $\boldsymbol{D}\boldsymbol{R}\textit{(t)=}\boldsymbol{A}\boldsymbol{R}\textit{(t)}$, $\boldsymbol{r}\textit{(0)=}\boldsymbol{I}$. As the solution of $\boldsymbol{D}\boldsymbol{r}\textit{(t)=}\boldsymbol{A}\boldsymbol{r}\textit{(t)}$, $\boldsymbol{r}\textit{(0)=}\boldsymbol{c}$, is $\boldsymbol{c}e^{\boldsymbol{A}t}$,[11] and the solution of $D\boldsymbol{R}(t) = \boldsymbol{A}\boldsymbol{R}(t)$, $\boldsymbol{R}(0) = \boldsymbol{I}$ is $e^{\boldsymbol{A}t}$, we have

$$\boldsymbol{y}\textit{(t)} = \boldsymbol{y}\textit{(0)}e^{\boldsymbol{A}t} + \int_0^t \boldsymbol{I}e^{\boldsymbol{A}(t-\theta)}\boldsymbol{w}(\theta)d\theta. \tag{48}$$

In the case of (43) let us note that that $\boldsymbol{w}\textit{(t)}$ is the sum of $\boldsymbol{B}\boldsymbol{z}\textit{(t)}$ and $\boldsymbol{u}\textit{(t)}$ and let us remember the property of the integral of a sum. Thus we can write

$$\boldsymbol{x}\textit{(t)} = \boldsymbol{x}\textit{(0)}e^{\boldsymbol{A}t} + \int_0^t e^{\boldsymbol{A}(t-\theta)}\boldsymbol{B}\boldsymbol{z}(\theta)d\theta + \int_0^t e^{\boldsymbol{A}(t-\theta)}\boldsymbol{u}(\theta)d\theta. \tag{49}$$

Given (44), we have $\boldsymbol{u}(\theta)d(\theta) = d\boldsymbol{\zeta}(\theta)$, so that the solution can be written as

$$\boldsymbol{x}\textit{(t)} = \boldsymbol{x}\textit{(0)}e^{\boldsymbol{A}t} + \int_0^t e^{\boldsymbol{A}(t-\theta)}\boldsymbol{B}\boldsymbol{z}(\theta)d\theta + \int_0^t e^{\boldsymbol{A}(t-\theta)}d\boldsymbol{\zeta}(\theta), \tag{50}$$

where the last integral on the right hand side is a stochastic integral.

Suppose now that the continuous variables are observed at equispaced intervals, say every δ time units, where δ is the length of the observation interval in terms of the basic time unit of the system. Evaluating (50) at time $t-\delta$ we have

[11] We recall that $e^{\boldsymbol{A}t}$ is defined as the matrix series $e^{\boldsymbol{A}t} = \sum_{j=0}^{\infty} \frac{\boldsymbol{A}^j t^j}{j!}$, $\boldsymbol{A}^0 = \boldsymbol{I}$, which converges uniformly in any finite interval, and the sum function is a continuous function of t for all finite t.

$$\boldsymbol{x}(t-\delta)=\boldsymbol{x}(0)e^{\boldsymbol{A}(t-\delta)}+\int_0^{t-\delta}e^{\boldsymbol{A}(t-\delta-\theta)}\boldsymbol{B}\boldsymbol{z}(\theta)d\theta+\int_0^{t-\delta}e^{\boldsymbol{A}(t-\delta-\theta)}d\boldsymbol{\zeta}(\theta). \quad (51)$$

By multiplying both members of (51) by $e^{\delta\boldsymbol{A}}$ and subtracting from (50) we obtain

$$\boldsymbol{x}(t)-e^{\delta\boldsymbol{A}}\boldsymbol{x}(t-\delta)=\int_0^{t}e^{\boldsymbol{A}(t-\theta)}\boldsymbol{B}\boldsymbol{z}(\theta)d\theta-\int_0^{t-\delta}e^{\boldsymbol{A}(t-\theta)}\boldsymbol{B}\boldsymbol{z}(\theta)d\theta+$$
$$+\int_0^{t}e^{\boldsymbol{A}(t-\theta)}d\boldsymbol{\zeta}(\theta)-\int_0^{t-\delta}e^{\boldsymbol{A}(t-\theta)}d\boldsymbol{\zeta}(\theta). \quad (52)$$

Now, since $\int_0^{t-\delta}...=-\int_{t-\delta}^{0}...$, and given the additive property of integrals, we get

$$\boldsymbol{x}(t)-e^{\delta\boldsymbol{A}}\boldsymbol{x}(t-\delta)=\int_{t-\delta}^{t}e^{\boldsymbol{A}(t-\theta)}\boldsymbol{B}\boldsymbol{z}(\theta)d\theta+\int_{t-\delta}^{t}e^{\boldsymbol{A}(t-\theta)}d\boldsymbol{\zeta}(\theta). \quad (53)$$

By suitable changes of variable[12] in the last two integrals we obtain

$$\boldsymbol{x}(t)=e^{\delta\boldsymbol{A}}\boldsymbol{x}(t-\delta)+\int_0^{\delta}e^{\boldsymbol{A}(\theta)}\boldsymbol{B}\boldsymbol{z}(t-\theta)d\theta+\int_0^{\delta}e^{\boldsymbol{A}(\theta)}d\boldsymbol{\zeta}(t-\theta). \quad (54)$$

Let $\boldsymbol{y}_\tau$ be the discrete observation of a continuous variable $\boldsymbol{y}(t)$ at time $\tau\delta$, τ being an integer, that is $\boldsymbol{y}_\tau=\boldsymbol{y}(\tau\delta)$. Setting $t=\tau\delta$ in (54) we have

[12]Consider, for example, the first integral on the right-hand side and make the substitution $\theta=t-s$. Then $\int_{t-\delta}^{t}e^{\boldsymbol{A}(t-\theta)}\boldsymbol{B}\boldsymbol{z}(\theta)d\theta=\int_\delta^0 e^{\boldsymbol{A}s}\boldsymbol{B}\boldsymbol{z}(t-s)\frac{d\theta}{ds}ds=-\int_\delta^0 e^{\boldsymbol{A}s}\boldsymbol{B}\boldsymbol{z}(t-s)ds=\int_0^\delta e^{\boldsymbol{A}s}\boldsymbol{B}\boldsymbol{z}(t-s)ds$. Given the arbitrariness of the denomination of the integration variable, we denominate s as θ (of course, this θ is not the same θ as before, but this fact has no influence on the evaluation of the integrals (54)).

$$\boldsymbol{x}_\tau = e^{\delta \boldsymbol{A}} \boldsymbol{x}_{\tau-1} + \boldsymbol{\psi}_\tau + \boldsymbol{\omega}_\tau, \tag{55}$$

where

$$\boldsymbol{\psi}_\tau \equiv \int_0^\delta e^{\boldsymbol{A}\theta} \boldsymbol{B} \boldsymbol{z}(\tau\delta - \theta) d\theta, \quad \boldsymbol{\omega}_\tau \equiv \int_0^\delta e^{\boldsymbol{A}\theta} d\boldsymbol{\zeta}(\tau\delta - \theta). \tag{56}$$

Equations (55) can be considered as a set of stochastic difference equations generating the discrete time data $\boldsymbol{x}_\tau$ over some sample period $\tau = 1, ..., T$. Since the observations generated by the differential system (43) will satisfy the exact discrete model (55) irrespective of the length of the interval between successive observations, the sampling properties of the differential system may be studied by considering the sampling properties of the exact discrete model.

We must now cope with two problems: the first is how to evaluate the integral $\boldsymbol{\psi}_\tau$, the second is how to deal with flow variables.

Concerning the evaluation of the integral $\boldsymbol{\psi}_\tau$, we can affirm that in general exogenous variables $\boldsymbol{z}(t)$ will not be simple integrable functions of time so that we cannot integrate out in (56) to obtain a model that can be estimated directly. Therefore an approximation to $\boldsymbol{\psi}_\tau$ is necessary. A satisfactory solution is provided by a quadratic approximation, namely we approximate $\boldsymbol{z}(\tau\delta - \theta)$ by a quadratic in θ for $0 \leq \theta < \delta$ and express the coefficients of this quadratic in terms of three consecutive observations $\boldsymbol{z}_{\tau-2}$, $\boldsymbol{z}_{\tau-1}$ and $\boldsymbol{z}_\tau$. Consider the vector equation $\boldsymbol{z}(\tau\delta - \theta) = \boldsymbol{a}\theta^2 + \boldsymbol{b}\theta + \boldsymbol{c}$. Then

$$\boldsymbol{z}_\tau = \boldsymbol{z}(\tau\delta) = \boldsymbol{z}(\tau\delta - \theta) \quad for \quad \theta = 0 \quad and \quad so \quad \boldsymbol{z}_\tau = \boldsymbol{c},$$

$$\boldsymbol{z}_{\tau-1} = \boldsymbol{z}(\tau\delta - \delta) = \boldsymbol{a}\delta^2 + \boldsymbol{b}\delta + \boldsymbol{c},$$

$$\boldsymbol{z}_{\tau-2} = \boldsymbol{z}(\tau\delta - 2\delta) = 4\boldsymbol{a}\delta^2 + 2\boldsymbol{b}\delta + \boldsymbol{c}. \tag{57}$$

By substituting $\boldsymbol{z}(\tau\delta - \theta) = \boldsymbol{a}\theta^2 + \boldsymbol{b}\theta + \boldsymbol{c}$ in the expression for $\boldsymbol{\psi}_\tau$ defined in (55), we have

$$\boldsymbol{\psi}_\tau = \int_0^\delta e^{\boldsymbol{A}\theta}\boldsymbol{B}(\boldsymbol{a}\theta^2 + \boldsymbol{b}\theta + \boldsymbol{c})d\theta. \tag{58}$$

If we integrate out in (58) we obtain[13]

$$\begin{aligned}\boldsymbol{\psi}_\tau = \int_0^\delta e^{\boldsymbol{A}\theta}\boldsymbol{B}\boldsymbol{a}\theta^2 d\theta + \int_0^\delta e^{\boldsymbol{A}\theta}\boldsymbol{B}\boldsymbol{b}\theta d\theta + \int_0^\delta e^{\boldsymbol{A}\theta}\boldsymbol{B}\boldsymbol{c} d\theta = \\ = [(\boldsymbol{A}^{-1}\delta^2 e^{\boldsymbol{A}\delta}\boldsymbol{B}\boldsymbol{a} - 2\boldsymbol{A}^{-2}e^{\boldsymbol{A}\delta}\boldsymbol{B}\boldsymbol{a}\delta + 2\boldsymbol{A}^{-3}e^{\boldsymbol{A}\delta}\boldsymbol{B}\boldsymbol{a}) - (2\boldsymbol{A}^{-3}\boldsymbol{B}\boldsymbol{a})] + \\ + [(\boldsymbol{A}^{-1}e^{\boldsymbol{A}\delta}\boldsymbol{B}\boldsymbol{b}\delta - \boldsymbol{A}^{-2}e^{\boldsymbol{A}\delta}\boldsymbol{B}\boldsymbol{b}) - (-\boldsymbol{A}^{-2}\boldsymbol{B}\boldsymbol{b})] + (\boldsymbol{A}^{-1}e^{\boldsymbol{A}\delta}\boldsymbol{B}\boldsymbol{c} - \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{c}).\end{aligned} \tag{59}$$

In the last equation we collect terms according to the negative powers of $\boldsymbol{A}$ and using (57) obtain

$$\begin{aligned}\boldsymbol{\psi}_\tau = \boldsymbol{A}^{-1}[e^{\boldsymbol{A}\delta}\boldsymbol{B}(\boldsymbol{a}\delta^2 + \boldsymbol{b}\delta + \boldsymbol{c}) - \boldsymbol{B}\boldsymbol{c}] - \boldsymbol{A}^{-2}[e^{\boldsymbol{A}\delta}\boldsymbol{B}(2\boldsymbol{a}\delta + \boldsymbol{b}) - \boldsymbol{B}\boldsymbol{b}] + 2\boldsymbol{A}^{-3}(e^{\boldsymbol{A}\delta} - \boldsymbol{I})\boldsymbol{B}\boldsymbol{a} = \\ = \boldsymbol{A}^{-1}[e^{\boldsymbol{A}\delta}\boldsymbol{B}\boldsymbol{z}_{\tau-1} - \boldsymbol{B}\boldsymbol{z}_\tau] - \boldsymbol{A}^{-2}[e^{\boldsymbol{A}\delta}\boldsymbol{B}(2\boldsymbol{a}\delta + \boldsymbol{b}) - \boldsymbol{B}\boldsymbol{b}] + 2\boldsymbol{A}^{-3}(e^{\boldsymbol{A}\delta} - \boldsymbol{I})\boldsymbol{B}\boldsymbol{a}.\end{aligned} \tag{60}$$

By direct substitution[14] we can check that

[13] From ordinary integral calculus remember that $\int e^{\alpha x} x dx = \alpha^{-2} e^{\alpha x}(\alpha x - 1)$ and that $\int e^{\alpha x} x^2 dx = \alpha^{-1} x^2 e^{\alpha x} - 2\alpha^{-1} \int e^{\alpha x} x dx = \alpha^{-1} x^2 e^{\alpha x} - 2\alpha^{-3} e^{\alpha x}(\alpha x - 1)$. These rules apply to matrix integrals as well.

[14] That is substituting in (61) the values of $\boldsymbol{z}_\tau$ etc. defined in (57). Apart from this check, the procedure to obtain (61) is the following. Let $\boldsymbol{\psi}_\tau = \boldsymbol{\alpha}\boldsymbol{B}\boldsymbol{z}_\tau + \boldsymbol{\beta}\boldsymbol{B}\boldsymbol{z}_{\tau-1} + \boldsymbol{\gamma}\boldsymbol{B}\boldsymbol{z}_{\tau-2}$, where $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are to be determined. Substituting $\boldsymbol{z}_\tau$ etc. from (57) and comparing the result with (60) we can determine $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$. The result is (61).

$$\boldsymbol{\psi}_\tau = \delta[[(\delta\boldsymbol{A})^{-3} + \frac{1}{2}(\delta\boldsymbol{A})^{-2}]e^{\boldsymbol{A}\delta} - (\delta\boldsymbol{A})^{-3} - \frac{3}{2}(\delta\boldsymbol{A})^{-2} - (\delta\boldsymbol{A})^{-1}]\boldsymbol{B}\boldsymbol{z}_\tau +$$
$$+\delta[[-2(\delta\boldsymbol{A})^{-3} + (\delta\boldsymbol{A})^{-1}]e^{\boldsymbol{A}\delta} + 2(\delta\boldsymbol{A})^{-3} + 2(\delta\boldsymbol{A})^{-2}]\boldsymbol{B}\boldsymbol{z}_{\tau-1} +$$
$$+\delta[[(\delta\boldsymbol{A})^{-3} - \frac{1}{2}(\delta\boldsymbol{A})^{-2}]e^{\boldsymbol{A}\delta} - (\delta\boldsymbol{A})^{-3} - \frac{1}{2}(\delta\boldsymbol{A})^{-2}]\boldsymbol{B}\boldsymbol{z}_{\tau-2} \qquad (61)$$

The bias involved in using this approximation is of $O(\delta^4)$ for small δ.[15] Of course (61) will be exact when the elements of $\boldsymbol{z}(t)$ are polynomials in t the degree of which is at most two. For estimation purpose we may consider $\delta = 1$ (this means setting the basic time unit equal to the length of the observation interval), hence $\tau = t$. When we come to estimate the model (55) in practice, it will often be necessary to use the derivatives of the expressions that appear in (61) with respect to the unknown elements of $\boldsymbol{A}$ and $\boldsymbol{B}$ in the computations. These derivatives give rise to quite complicated formulae which can lead to serious rounding errors. It is therefore useful to simplify the expressions before differentiating them; this simplification can be obtained by expanding $\boldsymbol{A}$ in a power series. To this purpose, let us set $\delta = 1$ in (61) and call $\boldsymbol{K}_1$, $\boldsymbol{K}_2$ and $\boldsymbol{K}_3$, in the order, the coefficient of $\boldsymbol{B}\boldsymbol{z}_t$, the coefficient of $\boldsymbol{B}\boldsymbol{z}_{t-1}$ and the coefficient of $\boldsymbol{B}\boldsymbol{z}_{t-2}$.

Since

$$e^{\boldsymbol{A}} = \sum_{k=0}^{\infty} \frac{\boldsymbol{A}^k}{k!} = \boldsymbol{I} + \boldsymbol{A} + \frac{1}{2}\boldsymbol{A}^2 + \frac{\boldsymbol{A}^3}{3!} + \frac{\boldsymbol{A}^4}{4!} + ... + \frac{\boldsymbol{A}^n}{n!} + ..., \qquad (62)$$

it follows that

$$\begin{aligned} \boldsymbol{A}^{-3}(e^{\boldsymbol{A}} - \boldsymbol{I}) &= \boldsymbol{A}^{-2} + \frac{1}{2}\boldsymbol{A}^{-1} + \frac{\boldsymbol{I}}{3!} + \frac{\boldsymbol{A}}{4!} + ... + \frac{\boldsymbol{A}^{n-3}}{n!} + ... = \\ &= \boldsymbol{A}^{-2} + \frac{1}{2}\boldsymbol{A}^{-1} + \boldsymbol{E}, \end{aligned} \qquad (63)$$

[15] The order of magnitude symbol O is used here in the following sense: given a function $f(\delta)$, $f(\delta) = O(\delta^r)$ if $\lim_{\delta\to 0}(\delta^{-r}|f(\delta)|)$ is finite.

where

$$\boldsymbol{E} \equiv \sum_{k=0}^{\infty} \frac{\boldsymbol{A}^k}{(k+3)!}. \tag{64}$$

Hence, as regards $\boldsymbol{K}_1$, we can write

$$\begin{aligned}
\boldsymbol{K}_1 &= (\boldsymbol{A}^{-3} + \frac{1}{2}\boldsymbol{A}^{-2})e^{\boldsymbol{A}} - \boldsymbol{A}^{-3} - \frac{3}{2}\boldsymbol{A}^{-2} - \boldsymbol{A}^{-1} \\
&= (\boldsymbol{I} + \frac{1}{2}\boldsymbol{A})\boldsymbol{A}^{-3}e^{\boldsymbol{A}} - (\boldsymbol{I} + \frac{1}{2}\boldsymbol{A})\boldsymbol{A}^{-3} - \boldsymbol{A}^{-2} - \boldsymbol{A}^{-1} \\
&= (\boldsymbol{I} + \frac{1}{2}\boldsymbol{A})\boldsymbol{A}^{-3}(e^{\boldsymbol{A}} - \boldsymbol{I}) - \boldsymbol{A}^{-2} - \boldsymbol{A}^{-1} \\
&= (\boldsymbol{I} + \frac{1}{2}\boldsymbol{A})(\boldsymbol{A}^{-2} + \frac{1}{2}\boldsymbol{A}^{-1} + \boldsymbol{E}) - \boldsymbol{A}^{2} - \boldsymbol{A}^{-1} \\
&= (\boldsymbol{I} + \frac{1}{2}\boldsymbol{A})\boldsymbol{E} + \frac{1}{4}\boldsymbol{I}.
\end{aligned} \tag{65}$$

Using (63), for $\boldsymbol{K}_2$ we can write

$$\begin{aligned}
\boldsymbol{K}_2 &= (-2\boldsymbol{A}^{-3} + \boldsymbol{A}^{-1})e^{\boldsymbol{A}} + 2\boldsymbol{A}^{-3} + 2\boldsymbol{A}^{-2} \\
&= (-2\boldsymbol{I} + \boldsymbol{A}^{2})\boldsymbol{A}^{-3}e^{\boldsymbol{A}} - (-2\boldsymbol{I} + \boldsymbol{A}^{2})\boldsymbol{A}^{-3} + 2\boldsymbol{A}^{-2} + \boldsymbol{A}^{-1} \\
&= (-2\boldsymbol{I} + \boldsymbol{A}^{2})\boldsymbol{A}^{-3}(e^{\boldsymbol{A}} - \boldsymbol{I}) + 2\boldsymbol{A}^{-2} + \boldsymbol{A}^{-1} \\
&= (-2\boldsymbol{I} + \boldsymbol{A}^{2})(\boldsymbol{A}^{-2} + \frac{1}{2}\boldsymbol{A}^{-1} + \boldsymbol{E}) + 2\boldsymbol{A}^{-2} + \boldsymbol{A}^{-1} \\
&= (-2\boldsymbol{I} + \boldsymbol{A}^{2})\boldsymbol{E} - 2\boldsymbol{A}^{-2} - \boldsymbol{A}^{-1} + \boldsymbol{I} + \frac{1}{2}\boldsymbol{A} + 2\boldsymbol{A}^{-2} + \boldsymbol{A}^{-1} \\
&= (-2\boldsymbol{I} + \boldsymbol{A}^{2})\boldsymbol{E} + \boldsymbol{I} + \frac{1}{2}\boldsymbol{A}.
\end{aligned} \tag{66}$$

Using (63), for $\boldsymbol{K}_3$ we can write

$$\begin{aligned}
\boldsymbol{K}_3 &= (\boldsymbol{A}^{-3} + \frac{1}{2}\boldsymbol{A}^{-2})e^{\boldsymbol{A}} - \boldsymbol{A}^{-3} + \frac{1}{2}\boldsymbol{A}^{-2} \\
&= (\boldsymbol{I} - \frac{1}{2}\boldsymbol{A})\boldsymbol{A}^{-3}e^{\boldsymbol{A}} - (\boldsymbol{I} - \frac{1}{2}\boldsymbol{A})\boldsymbol{A}^{-3} - \boldsymbol{A}^{-2} \\
&= (\boldsymbol{I} - \frac{1}{2}\boldsymbol{A})\boldsymbol{A}^{-3}(e^{\boldsymbol{A}} - \boldsymbol{I}) - \boldsymbol{A}^{-2} \\
&= (\boldsymbol{I} - \frac{1}{2}\boldsymbol{A})(\boldsymbol{A}^{-2} + \frac{1}{2}\boldsymbol{A}^{-1} + \boldsymbol{E}) - \boldsymbol{A}^{-2} \\
&= (\boldsymbol{I} - \frac{1}{2}\boldsymbol{A})\boldsymbol{E} + \boldsymbol{A}^{-2} + \frac{1}{2}\boldsymbol{A}^{-1} - \frac{1}{2}\boldsymbol{A}^{-1} - \frac{1}{4}\boldsymbol{I} - \boldsymbol{A}^{-2} \\
&= (\boldsymbol{I} - \frac{1}{2}\boldsymbol{A})\boldsymbol{E} - \frac{1}{4}\boldsymbol{I}.
\end{aligned} \tag{67}$$

With these transformations, system (55) can be written as

$$\boldsymbol{x}_t = e^{\boldsymbol{A}}\boldsymbol{x}_{t-1} + \sum_{i=1}^{3} \boldsymbol{C}_i \boldsymbol{z}_{t-i+1} + \boldsymbol{\omega}_t, \tag{68}$$

where the coefficients $\boldsymbol{C}_i$ are defined as the product $\boldsymbol{K}_i\boldsymbol{B}$, that is

$$\boldsymbol{C}_1 \equiv [(\boldsymbol{I} + \frac{1}{2}\boldsymbol{A})\boldsymbol{E} + \frac{1}{4}\boldsymbol{I}]\boldsymbol{B}, \tag{69}$$

$$\boldsymbol{C}_2 \equiv [(-2\boldsymbol{I} + \boldsymbol{A}^2)\boldsymbol{E} + \boldsymbol{I} + \frac{1}{2}\boldsymbol{A}]\boldsymbol{B}, \tag{70}$$

$$\boldsymbol{C}_3 \equiv [(\boldsymbol{I} - \frac{1}{2}\boldsymbol{A})\boldsymbol{E} - \frac{1}{4}\boldsymbol{I}]\boldsymbol{B}, \tag{71}$$

and E is defined in (64).

Now let us cope with the other problem mentioned before and concerning models containing flow variables. We will see that if the original model contains some flow variables, the data must be pre-whitened to eliminate the autocorrelation introduced with the integration, and that this is obtained by using a certain transformation. We will also see that, before using this transformation, in a model of this kind, instantaneous variables must be previously averaged using a certain operator and that variables which are the rate of change of some instantaneously measurable variables must be previously replaced by the first difference of the instantaneous variables from which they derive.[16]

The exact discrete analogue has been derived under the assumption that all variables are measurable at a point in time. Although this is true for the instantaneous variables like wages, prices, interest rates and stocks, many economic variables are flow variables such as consumption, income, etc. or variables which are the rate of change of some instantaneously measurable variables, such as the rate of change of the money supply, the rate of change of the capital stock (net investment), etc. These variables (which we term flow variables in the broad sense) are not measurable at a point in time and observations on them are usually integrals over some given interval. If we let $y(t)$ be a generic flow variable, the integral

$$y^0(t) = \frac{1}{\delta} \int_0^\delta y(t-\theta)d\theta \qquad (72)$$

is measurable. We denote by y^0_τ an observation of $y^0(t)$, that is a flow over the interval $(\tau\delta - \delta, \tau\delta)$ where δ is, as before, the length of the observation interval; actually y^0_τ is what we observe in reality. Therefore, a continuous model containing flow variables needs to be integrated over the observation interval in order to produce a model defined in terms of variables that are

[16]In other words, when flow variables are present in the model, it must be integrated over the observation interval in order to obtain a model defined in terms of measurable variables. But in this way we introduce autocorrelation in the new disturbance term, whereas the disturbances in the original model were exempt from autocorrelation by assumption. The new disturbances, however, have the same autocorrelation properties of a certain moving average process.

measurable. If we integral (54)[17] we obtain

$$\frac{1}{\delta}\int_{t-\delta}^{t}\boldsymbol{x}(\theta)d\theta = e^{\delta\boldsymbol{A}}\frac{1}{\delta}\int_{t-\delta}^{t}\boldsymbol{x}(\theta-\delta)d\theta +$$
$$+\frac{1}{\delta}\int_{t-\delta}^{t}\int_{0}^{\delta}e^{\boldsymbol{A}\theta}\boldsymbol{B}\boldsymbol{z}(s-\theta)d\theta ds + \frac{1}{\delta}\int_{t-\delta}^{t}\int_{0}^{\delta}e^{\boldsymbol{A}\theta}d\boldsymbol{\zeta}(s-\theta)ds. \tag{73}$$

With the property of the change in the order of integration in multiple integrals we can write

$$\frac{1}{\delta}\int_{t-\delta}^{t}\boldsymbol{x}(\theta)d\theta = e^{\delta\boldsymbol{A}}\frac{1}{\delta}\int_{t-\delta}^{t}\boldsymbol{x}(\theta-\delta)d\theta +$$
$$+\int_{0}^{\delta}e^{\boldsymbol{A}\theta}\frac{1}{\delta}\int_{t-\delta}^{t}\boldsymbol{B}\boldsymbol{z}(s-\theta)dsd\theta + \frac{1}{\delta}\int_{t-\delta}^{t}\int_{0}^{\delta}e^{\boldsymbol{A}\theta}d\boldsymbol{\zeta}(s-\theta)ds.^{18} \tag{74}$$

Thus the integral of a flow variable can be replaced by the observed value $y^0(t)$ as defined in (72), and y can be either an endogenous x or an exogenous z variable. Now, apart from the particular case of a model containing only flow variables, the integration carried out in (73) and in (74) has also transformed the instantaneous variables into integrals and these we must deal with. In fact, if we now let $y(t)$ denote a generic instantaneous variable, the integral $\frac{1}{\delta}\int_{t-\delta}^{t}y(\theta)d\theta$ is not observable and has to be evaluated using, for example, the trapezoidal rule. If δ is small, an acceptable approximation is

$$\int_{t-\delta}^{t}y(\theta)d\theta \simeq \frac{\delta}{2}[(y(t)+y(t-\delta)], \tag{75}$$

whence, if we set $t=\tau\delta$,

[17] Consider that: $\boldsymbol{x}(t) = \frac{1}{\delta}\int_0^\delta \boldsymbol{x}(t-\theta)d\theta = \frac{1}{\delta}\int_{t-\delta}^{t}\boldsymbol{x}(\theta)d\theta$, that $e^{\delta\boldsymbol{A}}\boldsymbol{x}(t-\delta) = e^{\delta\boldsymbol{A}}\frac{1}{\delta}\int_0^\delta \boldsymbol{x}(t-\delta-\theta)d\theta = e^{\delta\boldsymbol{A}}\frac{1}{\delta}\int_{t-\delta}^{t}(\theta-\delta)d\theta$, and that $\int_0^\delta e^{\boldsymbol{A}\theta}\boldsymbol{B}\boldsymbol{z}(t-\theta)d\theta = \frac{1}{\delta}\int_0^\delta\int_0^\delta e^{\boldsymbol{A}\theta}\boldsymbol{B}\boldsymbol{z}(t-\theta-s)d\theta ds = \frac{1}{\delta}\int_{t-\delta}^{t}\int_0^\delta e^{\boldsymbol{A}\theta}\boldsymbol{B}\boldsymbol{z}(\delta-\theta)d\theta ds$.

[18] Note that $\int_{t-\delta}^{t}\boldsymbol{y}(\theta)d\theta = \int_0^\delta \boldsymbol{y}(t-\theta)d\theta$ and that $\int_0^\delta \boldsymbol{y}(t-\theta)d\theta = \int_{t-\delta}^{t}\boldsymbol{y}(\delta)d\delta$.

$$\frac{1}{\delta}\int_{t-\delta}^{t} y(\theta)d\theta \simeq \frac{1}{2}[(y(\tau\delta)+y(\tau\delta-\delta)], \tag{76}$$

so that an observation of this integral is approximately given by[19]

$$y_\tau^0 = My_\tau, \quad M \equiv \frac{1}{2}(1+L), \tag{77}$$

where L is the lag operator. It should be pointed out that when we have to deal with a variable that in principle is an instantaneous variable but that has actually been observed or defined as an average of the values over the observation period (e.g. an annual price index defined as the average of the monthly indexes) or generated in such a way as to give observations equivalent to period averages (e.g. an implicit GNP deflator), then the variable that we observe must be treated as the integral of the continuous instantaneous variable. In other words, in these cases (72) holds and not (77). It may also happen that the model contains second order equations that determine stock variables. In this case, as we know, we introduce additional variables so as to reduce the model to a first-order system. These variables will be the instantaneous rates of change of the stock variables, and observations on them can be derived from the point observations on the stock. Let $y_1(t) = Dy(t)$ be the new variable; then we have

$$\frac{1}{\delta}\int_{t-\delta}^{t} y_1(\theta)d\theta = \frac{1}{\delta}[(y(t)+y(t-\delta)], \tag{78}$$

and observations on $y_1(\tau\delta)$ are given by [20]

$$y_{1\tau}^0 = \Delta y_\tau, \quad \Delta \equiv \frac{1}{\delta}(1-L). \tag{79}$$

[19]In fact we have that: $y_\tau^0 = My_\tau = \frac{1}{2}(1+L)y_\tau = \frac{1}{2}y_\tau + \frac{1}{2}Ly_\tau = \frac{1}{2}y_\tau + \frac{1}{2}y_{\tau-1} = \frac{1}{2}y(\tau\delta) + \frac{1}{2}y((\tau-1)\delta) = \frac{1}{2}y(\tau\delta)\frac{1}{2}y(\tau\delta-\delta)$.

[20]In fact we have that: $y_{1\tau}^0 = \Delta y_\tau = \frac{1}{\delta}(1-L)y_\tau = \frac{1}{\delta}(y_\tau - Ly_\tau) = \frac{1}{\delta}(y_\tau - y_{\tau-1}) = \frac{1}{\delta}[y(\tau\delta) - y((\tau-1)\delta)] = \frac{1}{\delta}[y(t) - y(\tau-\delta)]$.

Thus the exact discrete analogue of a continuous mixed stock-flow model is given by

$$\boldsymbol{x}_\tau^0 = e^{\delta \boldsymbol{A}} \boldsymbol{x}_{\tau-1}^0 + \int_0^\delta e^{\boldsymbol{A}\theta} \boldsymbol{B} \boldsymbol{z}^0(\tau\delta - \theta) d\theta + \int_{\tau\delta-\delta}^{\tau\delta} e^{\boldsymbol{A}\theta} d\boldsymbol{\zeta}(s-\theta) ds, \tag{80}$$

where the variables with the subscript zero are replaced by (72), (77), or (79) depending on the case. We must next point out that the integration carried out in order to obtain measurable variables implies that the disturbances in (74) and so in (80) are no longer serially uncorrelated. It is possible, however, to derive an approximation to the process of formation of these disturbances that is independent of the parameters of the model and which allows a pre-whitening of the data. If we denote by $\boldsymbol{\xi}(t)$ the disturbance vector, this approximation is[21]

$$\boldsymbol{\xi}(t) \simeq (1 + 0.268L)\boldsymbol{\epsilon}(t), \tag{81}$$

where $\boldsymbol{\epsilon}(t)$ is a serially uncorrelated random disturbance. As the moving average process (81) is independent of the parameters of the model, the model can be transformed by using the inverse of this process to obtain a model with serially uncorrelated disturbances. We have (as we are in discrete time, we use t as subscript)

$$\boldsymbol{\epsilon}_t = (1 + 0.268L)^{-1}\boldsymbol{\xi}_t. \tag{82}$$

If we expand $(1 + 0.268L)^{-1}$ in a Taylor series[22] and truncate after the

[21] The approximation is due to the fact that the factor $e^{\boldsymbol{A}\theta}$ is neglected in the disturbance term in (74) and so in (80); assuming that the observation interval is small this is acceptable because the elements of $\delta\boldsymbol{A}$ will be small.

[22] As the operator L can be treated as an algebraic quantity, we use the power series expression $(1+x)^\alpha = 1 + \alpha x + \frac{\alpha(\alpha-1)}{2!}x^2 + \frac{\alpha(\alpha-1)(\alpha-2)}{3!}x^3 + \ldots = \sum_{n=0}^{\infty} \frac{\alpha(\alpha-1)\ldots(\alpha-n+1)}{n!}x^n$, which, for $\alpha = -1$, becomes $1 - x + x^2 - x^3 + \ldots$. In our case, x=0.268L.

third order term which is sufficiently small for all practical purposes, we obtain

$$\begin{aligned}\boldsymbol{\epsilon}_t &\simeq (1-0.268L+0.268^2L^2-0.268^3L^3)\boldsymbol{\xi}_t = \\ &= \boldsymbol{\xi}_t - 0.268\boldsymbol{\xi}_{t-1} + 0.268^2\boldsymbol{\xi}_{t-2} - 0.268^3\boldsymbol{\xi}_{t-3}.\end{aligned} \tag{83}$$

Therefore if we transform the variable $\boldsymbol{x}^0$ and $\boldsymbol{z}^0$ in (80) by means of the same process, that is if (setting $\delta = 1$ and so $\tau = t$) we let

$$\boldsymbol{x}_t^* = \boldsymbol{x}_t^0 - 0.268\boldsymbol{x}_{t-1}^0 + 0.268^2\boldsymbol{x}_{t-2}^0 - 0.268^3\boldsymbol{x}_{t-3}^0, \tag{84}$$

and similarly for $\boldsymbol{z}$ (remembering that the disturbances term in (80) is $\boldsymbol{\xi}_t$), we obtain the following model with serially uncorrelated disturbances

$$\boldsymbol{x}_t^* = e^{\boldsymbol{A}}\boldsymbol{x}_{t-1}^* + \int_0^1 e^{\boldsymbol{A}\theta}\boldsymbol{B}\boldsymbol{z}^*(t-\theta)d\theta + \boldsymbol{\epsilon}_t, \tag{85}$$

where the integral can be approximate as we so previously in (58). The parameters of this model - or of model (68), if no flow variables are present - can by estimated by using FIML method.

However, as this estimation can be time consuming and therefore expensive, especially if there are general restrictions on the elements of $\boldsymbol{A}$ and $\boldsymbol{B}$, it could be convenient to consider a discrete approximation to system (43). Concerning the approximate discrete analogue, let us integrate the system (43) over the interval $(t-\delta, t)$ and utilize the following approximations[23]

$$\frac{1}{\delta}\int_0^\delta \boldsymbol{D}\boldsymbol{y}(\tau\delta-\theta)d\theta = \boldsymbol{\Delta}\boldsymbol{y}_\tau, \quad \int_0^\delta \boldsymbol{y}(\tau\delta-\theta)d\theta = \boldsymbol{M}\boldsymbol{y}_\tau, \tag{86}$$

[23]Since the integral of the derivative of a function is the function itself, the first of the two relationships is exact; on the contrary, the second is approximate, as clarified in another context when dealing with (76).

where the operators $\boldsymbol{\Delta}$ and $\boldsymbol{M}$ have already been defined in (77) and (79), $\boldsymbol{y}$ is a generic vector of continuous variables and $\boldsymbol{y}_\tau$ is the discrete observation of $\boldsymbol{y}(\tau\delta)$ at time $t = \tau\delta$. Thus, as $\boldsymbol{MAy}_\tau = \boldsymbol{AMy}_\tau$, we have

$$\boldsymbol{\Delta x}_\tau = \boldsymbol{AMx}_\tau + \boldsymbol{BMz}_\tau + \boldsymbol{\eta}_\tau, \tag{87}$$

where $\boldsymbol{\eta}_\tau$ is a vector of disturbances that depends on $\boldsymbol{u}$ and on the errors of approximation. As we already saw in the exact discrete analogue, the disturbances $\boldsymbol{\eta}_\tau$ will be serially uncorrelated if all the variables are measured at a point in time. In the opposite case, in particular in mixed stock-flow models, model (43) must be integrated twice, once to provide measurable variables and once to obtain the approximate discrete analogue. By performing these integrations we obtain

$$\boldsymbol{\Delta x}^0_\tau = \boldsymbol{AMx}^0_\tau + \boldsymbol{BMz}^0_\tau + \boldsymbol{v}_\tau, \tag{88}$$

where the variables with a superscript zero are as defined in (72) and in (77) respectively for flow and stock variables. Also in this case the disturbances $\boldsymbol{v}_\tau$ will no longer be serially uncorrelated, but the same transformation described in the exact discrete analogue - see equation (84) - allows us to obtain a model with serially uncorrelated disturbances. The approximate discrete model (88) - after the just mentioned transformation of the variables, if it is the case - can be estimated using a FIML estimator with any a priori constraint on the parameters.

Let us now examine the bias introduced by the discrete approximation.

To this purpose we compare the coefficients of the approximate discrete model (87) with those of the exact discrete model (55). Given the definitions of the operators $\boldsymbol{\Delta}$ and $\boldsymbol{M}$, model (87) can be rewritten with simple passages as[24]

[24] In fact consider that $\Delta \boldsymbol{x}_\tau = \boldsymbol{AMx}_\tau$, hence $\frac{1}{\delta}(1-L)\boldsymbol{x}_\tau = \boldsymbol{A}\frac{1}{2}(1+L)\boldsymbol{x}_\tau = \boldsymbol{A}\frac{1}{2}(\boldsymbol{x}_\tau + \boldsymbol{x}_{\tau-1})$, so $\frac{1}{\delta}\boldsymbol{x}_\tau - \frac{1}{\delta}\boldsymbol{x}_{\tau-1} = \frac{1}{2}\boldsymbol{Ax}_\tau - \frac{1}{2}\boldsymbol{Ax}_{\tau-1}$, hence $\frac{1}{\delta}\boldsymbol{x}_\tau - \frac{1}{2}\boldsymbol{Ax}_\tau = \frac{1}{\delta}\boldsymbol{x}_{\tau-1} - \frac{1}{2}\boldsymbol{Ax}_{\tau-1}$, and so $\frac{1}{\delta}(\boldsymbol{I} - \frac{1}{2}\delta\boldsymbol{A})\boldsymbol{x}_\tau = \frac{1}{\delta}(\boldsymbol{I} + \frac{1}{2}\delta\boldsymbol{A})\boldsymbol{x}_{\tau-1}$.

$$\frac{1}{\delta}(\boldsymbol{I}-\frac{1}{2}\delta\boldsymbol{A})\boldsymbol{x}_{\tau}=\frac{1}{\delta}(\boldsymbol{I}+\frac{1}{2}\delta\boldsymbol{A})\boldsymbol{x}_{\tau-1}+\boldsymbol{BM}\boldsymbol{z}_{\tau}+\boldsymbol{\eta}_{\tau}. \tag{89}$$

If we premultiply both members by $\delta(\boldsymbol{I}-\frac{1}{2}\delta\boldsymbol{A})^{-1}$ and compare the matrix of coefficients of the lagged endogenous variables with the corresponding matrix, $e^{\delta\boldsymbol{A}}$, of the exact discrete model (55), we obtain the bias by difference

$$\begin{aligned}(\boldsymbol{I}-\frac{1}{2}\delta\boldsymbol{A})^{-1}(\boldsymbol{I}+\frac{1}{2}\delta\boldsymbol{A})-e^{\delta\boldsymbol{A}}=\\
=(\boldsymbol{I}+\frac{1}{2}\delta\boldsymbol{A}+\frac{1}{4}\delta^2\boldsymbol{A}^2+\frac{1}{8}\delta^3\boldsymbol{A}^3+...)(\boldsymbol{I}+\frac{1}{2}\delta\boldsymbol{A})+\\
-(\boldsymbol{I}+\delta\boldsymbol{A}+\frac{1}{2}\delta^2\boldsymbol{A}^2+\frac{1}{3!}\delta^3\boldsymbol{A}^3+...).\end{aligned} \tag{90}$$

Carrying out the multiplication on the right-hand side and collecting terms we have

$$\begin{aligned}(\boldsymbol{I}-\frac{1}{2}\delta\boldsymbol{A})^{-1}(\boldsymbol{I}+\frac{1}{2}\delta\boldsymbol{A})-e^{\delta\boldsymbol{A}}=(\boldsymbol{I}+\delta\boldsymbol{A}+\frac{1}{2}\delta^2\boldsymbol{A}^2+\frac{1}{4}\delta^3\boldsymbol{A}^3+...)+\\
-(\boldsymbol{I}+\delta\boldsymbol{A}+\frac{1}{2}\delta^2\boldsymbol{A}^2+\frac{1}{3!}\delta^3\boldsymbol{A}^3+...)=\frac{1}{12}\delta^3\boldsymbol{A}^3+\boldsymbol{O}(\delta^4),\end{aligned} \tag{91}$$

so that the approximation errors decrease with the cube of the observation interval. A similar result holds for the coefficients of the exogenous variables.[25]

Though the exact discrete model gives more precise estimates - with smaller asymptotic standard errors - of the parameters of the continuous model, with the approximate discrete model these estimates could be fairly satisfactory. However, even if one wants to obtain more precise estimates by

[25]It was shown through Monte Carlo studies (Phillips, 1974) that the bias is small also for quarterly models.

using the exact discrete model, the approximate discrete analogue could previously be used to determine the structure of the model, and to provide initial values for the parameters to be used as a starting point in the estimation with the exact discrete analogue. Finally, the determination of the structure of the model, using initially the approximate discrete analogue, usually requires a preliminary screening of several hypotheses and so the estimation of several preliminary versions of the model, which would be too expensive to carry out by using the exact discrete analogue.

CHAPTER 2: A continuous time model on economic growth and technology diffusion

2.1 Introduction

This chapter shows a model suitable for econometric purposes and based on economic theory, particularly the theory of the firm. We estimate a continuous time model of endogenous growth and technology diffusion. We explore the role of services, research activity and ICT - Information and Communication Technology - spending in technology diffusion. In studying the determinants of technology diffusion we also refer to the role of distance as it evolves over time. The model is estimated on several European countries and United States.

We have found that services, researchers and ICTs represent an important channel for technology diffusion and that effect of distance on such diffusion decreases over time.

2.2 The model

2.2.1 Conceptual framework

Economists have increasingly looked into the issue of integrating the accumulation of technology into growth models. In particular, we have to mention the seminal contributions of Romer (1990), Grossman and Helpman (1991) and Aghion and Howitt (1992). Other studies (Eaton and Kortum, 1996, 1997, 1999; Keller, 2002; Peri, 2004) have explicitly modeled and estimated the process of technology diffusion, or investigated the role of services, research activity and ICT in technologically driven growth (Francois, 1990; Mun and Nadiri, 2002). Instead, few studies have investigated the role of these channels in the diffusion of technology by directly modeling growth through endogenization of both the diffusion of technology and some of its channels (Guerrieri et al., 2005; Maggi, 2009).

In this study, we directly model output growth by endogenizing both the diffusion of technology and all its channels taken into account. These channels are given by services, researchers and ICTs. In particular, we analyse how they interact with technology accumulation and diffusion on the basis of their discussed key role in the innovation process.

Investigating the role of services as channel for technology diffusion represents a novel feature. In fact, literature has so far devoted little attention to the tertiary sector as driver of technology accumulation, and empirical analyses have almost entirely focused on manufacturing sector when studying the interaction between technology diffusion and growth.

Analysing the diffusion of technology, we also introduce space dimension. Accumulation of technology in each country depends on the extent this country can absorb technology produced in other countries. In fact, in our model, we use patent citations (a bilateral variable) to measure technology.[26] Nevertheless, the amount of foreign produced technology that can be used domestically is also limited by two factors: distance and absorption capacity in the receiving countries. These two factors are taken into account. In particular, we assume that the contribution of foreign technology to domestic technology accumulation grows as a negative function of distance from the countries from which flows of technology are acquired. Moreover, the impact of distance is allowed to vary over time to the extent that technological progress brings forward a reduction in the cost of technology diffusion. This last factor helps us to investigate the ability of receiving countries to use imported technology.

The analysis is performed referring to a context of EU integration.

2.2.2 The model equations and explanation of technology diffusion

The model is derived from maximizing an inter-temporal profit function, taking into account the costs of changing employment, investments and technology and assuming the production function $f(\cdot)$ is Cobb-Douglas. In this study we will consider n foreign countries.

Let L be the labour, K the capital stock, T the technology, Pat_{ij} the patent citations from country i to country j, I the investments and z the variation of labour. Let c, h and q be the costs of adjustment of z, I and the variation of technology, respectively; ρ is the rate of time preference and

[26] In the model besides the patent citations,we also have the distance as bilateral variable.

w the wages. For simplicity's sake we omit the residual terms and refer to chapter 1, paragraph 1.4, for an analysis of the stochastic properties of residuals in continuous time.

The discounted present value of the firm is

$$\max_{z,I,\sum_{i=1}^{n} Pat_{ij}} \int_t^{\infty} e^{-\rho s}[f(L,K,T) - wL - \frac{c}{2}z^2 - (1+\frac{h}{2}I)I - (1+\frac{q}{2}\sum_{i=1}^{n} Pat_{ij}) * \\ * \sum_{i=1}^{n} Pat_{ij}]\, ds, \quad j = 1,...,n$$

subject to

$$\begin{aligned} \dot{L} &= z, \\ \dot{K} &= I, \\ \dot{T} &= \sum_{i=1}^{n} Pat_{ij}, \quad j = 1,...,n. \end{aligned} \tag{92}$$

We also consider

$$\dot{\sum_{i=1}^{n} Pat_{ij}} = \delta[\beta_1 + \beta_2 T + (a+bt)\frac{dist_{.j}}{n} - \sum_{i=1}^{n} Pat_{ij}], \quad j = 1,...,n. \tag{93}$$

It was assumed that the derivatives of employment, capital and technology can be changed by the firms. As we can see observing the second constraint, I is considered as net investment.

It is worth to deeper analysing the third constraint and explaining how the domestic stock of technology in each country is given by the cumulated flow of patents obtained both through production and diffusion.

As technology variable we use patent citations; focusing on a single patent, we can consider its made citations and its received citations. In the first case we refer to the citations made by this patent to the other ones, in the second case we refer to the citations received by this patent from the other patents. A citation received between two countries indicates a transfer of technology. Hence, since Pat_{ij} represents the patent citations between countries, it captures the diffusion of technology between two countries and determine the accumulation of the domestic stock of technology. Hence we can indifferently refer to Pat_{ij} as bilateral flows of patents or bilateral exchange of technology. In other words, flows of patent citations (Pat) measure the change in the accumulation of the stock of technology. Bilateral flows of patent citations (Pat_{ij}) capture the diffusion of technology between two countries. Citations to country j occur when a patent whose inventor is resident in another country i mentions another patent, whose inventor is obviously original of country i, for the contribution it gives to the mentioning invention.

The stock of technology in each country evolves over time from *t-1* to t, given the initial condition, as follows

$$T_{tj} = T_{t-1j} + \sum_{i=1}^{n} Pat_{ijt}; \tag{94}$$

where the first subscript of Pat indicates the sender or cited country and the second subscript the recipient or citing country. The process starts at $t-1$ while Pat_{jjt} indicates the domestic accumulation of patents at time t and Pat_{ijt} indicates the amount of technology produced in country i that is actually received by country j at time t. In other words, technology accumulation in each country is disaggregated in two elements: technology accumulated domestically (domestic technology accumulation component) and the amount of technology accumulated in each of the other countries that is transferred to the recipient country through diffusion (imported technology accumulation component).

With this specification in mind we can now set out the differential equations of the model representing the third constraint. We have

$$T_{tj} = T_{0j} + \int_0^t (\sum_{i=1}^{n} Pat_{ijt})dt, \tag{95}$$

or

$$DT_j = \sum_{i=1}^{n} Pat_{ij}. \tag{96}$$

Also equation (93) requires further explanations. This is a partial adjustment equation[27] where $\beta_1 + \beta_2 T + (a + bt)\frac{dist_{.j}}{n}$ can be interpreted as the desired or potential value of $\sum_{i=1}^{n} Pat_{ij}$, δ is the speed of adjustment[28], $dist_{.j}$ stays for the distance between the countries i and j, and β_1, β_2, a and b are some parameters. Its derivation, step by step, can allow us to better understand its meaning.

In order to have an as much as possible disaggregate model, we have started from the following expression

$$\dot{Pat}_{ij} = \delta[S_{ij} + R_{ij} + ICT_{ij} + (a + bt)dist_{ij} - Pat_{ij}], \quad i, j = 1, ..., n, \tag{97}$$

where S are the services, R the researchers and ICT is the acronym of Information and Communication Technology. Clearly, if the last equation is true respect to a single country i, it will be also true from an aggregate point of view. Hence, considering the summation with respect to i, we have

[27]See chapter 1, paragraph 1.2 for more details on this kind of equations.

[28]As seen in chapter 1 - paragraph 1.2, in continuous time the speed of adjustment can be interpreted in terms of the mean time lag, as its reciprocal represents the time required for about 63% of the difference between the observed and the desired variables to be eliminated.

$$\sum_{i=1}^{n} \dot{Pat}_{ij} = \delta[\sum_{i=1}^{n} S_{ij} + \sum_{i=1}^{n} R_{ij} + \sum_{i=1}^{n} ICT_{ij} + (a+bt)\frac{\sum_{i=1}^{n} dist_{ij}}{n} - \sum_{i=1}^{n} Pat_{ij}],$$
$$j = 1, ..., n. \tag{98}$$

This last relation, if expressed with the aim to underlie the dependence of the services, researchers and ICTs on the technology, becomes

$$\sum_{i=1}^{n} \dot{Pat}_{ij} = \delta\{g[\sum_{i=1}^{n} S_{ij}(T_{ij}), \sum_{i=1}^{n} R_{ij}(T_{ij}), \sum_{i=1}^{n} ICT_{ij}(T_{ij})] + (a+bt)\frac{\sum_{i=1}^{n} dist_{ij}}{n} +$$
$$- \sum_{i=1}^{n} Pat_{ij}\}, \quad j = 1, ..., n, \tag{99}$$

where

$$S_{ij}(T_{ij}) = \gamma_1' + \gamma_2' T_{ij}, \quad i, j = 1, ..., n, \tag{100}$$
$$R_{ij}(T_{ij}) = \gamma_1'' + \gamma_2'' T_{ij}, \quad i, j = 1, ..., n, \tag{101}$$
$$ICT_{ij}(T_{ij}) = \gamma_1''' + \gamma_2''' T_{ij}, \quad i, j = 1, ..., n, \tag{102}$$

and $g(\cdot)$ is a function for the technology.

Expressing the equation explicitly with respect to the technology, we have our relation reported below

$$\sum_{i=1}^{n} \dot{Pat}_{ij} = \delta[\beta_1 + \beta_2 T + (a+bt)\frac{dist_{.j}}{n} - \sum_{i=1}^{n} Pat_{ij}], \quad j = 1, ..., n, \tag{103}$$

where, obviously,

$$\beta_1 = n\gamma_1' + n\gamma_1'' + n\gamma_1''' \tag{104}$$

and

$$\beta_2 = n\gamma_2' + n\gamma_2'' + n\gamma_2''',\text{[29]} \tag{105}$$

and where we have written T without the subscript ij just to be coherent with the notation given to the other variables appearing in the optimization problem.[30] In particular, as all the variables are given by the contribution not only of the country under consideration but also by the other countries, all of them should be written with the subscript ij, or at least with the subscript j, showing that j is the country under consideration and that has received its contribution respect to a particular variable by the other countries and also by itself. Nevertheless, for simplicity in the notation and for routinely, we drop the subscription.

The theoretical explanation of these relations is the following: the stock of technology produced in country i and acquired by country j, that is T_{ij}, depends on its contribution to country j's growth; the more the importance of the technology for country j's income, the more j invests in country i's technology.[31] In addition, the stock of technology causes the need for country j to own a certain amount of services, researchers and ICTs functional to this technology. For this reason, these variables are also reported with

[29]It is worth to noticing that the breaking down in such a way of the coefficients β_1 and β_2, which represent the impact of the stock of technology on the variation of patent flows, is very interesting, in fact, it allows us to understand how much of this effect is given by the services, researchers and ICTs.

[30]Clearly, in order to evaluate the impact of the single elements of the patent summation and hence establish how much the stock of technology of the generic country j depends on the flow of technology from another given country i - and also on its domestic production of technology -, we have to write this summation with respect to every single country i.

[31]Let the lecturer note that among the sender countries indicated with i there is also the country j.

the subscript ij: the services, researchers and ICTs from country i that are due to the demand of i's technology, that is $\dot{T}$ or $\sum_{i=1}^{n} Pat_{ij}$. For example, $S_{ij}(T)$ represents the contribution of the services from country i to country j through country i's technology, in other words services facilitate the use of technology stemming from country i. Similarly, R_{ij} is the contribution of country's i researchers towards country's j for the technology acquired from country i. The same reasoning holds for ICTs. Now, the contributions of the technology which impacts on S, R and ICT determine the need for the patents. Hence, S, R and ICT are functional to the patents and affected by the technology.

Finally, the introduction of the mean distance certainly contributes to explain the decision for a given country to receive a certain amount of patents coming from another country.[32] However, thanks to the lower costs of transferring technology and information across space, this decision is always less influenced by the distance.[33] Hence, in the evaluation of the transmission of technology, we consider not only the impact of the geographical distance from the country from which patents are received on the amount of patents, but we also analyse how this impact decreases over time. This is made formalizing the effect of distance by a direct impact a and by an indirect one after a change in the time given by b. In particular, conscious that distance represents always less an obstacle in the relationships among countries, its introduction in the model is just a way to test this fact.

To summarize, for each country of the panel data, we have the following endogenous variables: $T, K, L, S, R, ICT, Pat_{ij}$; and the following exogenous variables: $dist_{ij}, t$.

The Hamiltonian of our dynamic optimization problem is

[32] Obviously, having introduced the distance in order to analyse the flows of technology between countries, it seems more appropriate consider the mean distance of a specific country respect to the other ones instead of considering just the summation of such distances.

[33] However, as Peri (2004) notes, time could have a negative impact to the extent that the value of innovation in a patent decreases over time with obsolescence.

$$H = e^{-\rho t}[f(L,K,T) - wL - \frac{c}{2}z^2 - (1+\frac{h}{2}I)I - (1+\frac{q}{2}\sum_{i=1}^{n} Pat_{ij})\sum_{i=1}^{n} Pat_{ij}]+$$

$$+ v_1 z + v_2 I + v_3 \sum_{i=1}^{n} Pat_{ij}, \quad j = 1,...,n, \tag{106}$$

where

$$v_m = \mu_m e^{-\rho t}, \qquad for \quad m = 1,...,3, \tag{107}$$

and

$$\dot{v_m} = \dot{\mu_m} e^{-\rho t} - \rho \mu_m e^{-\rho t}, \quad for \quad m = 1,...,3, \tag{108}$$

are the costate variables.

The first order conditions are

$$\frac{\partial H}{\partial v_1} = \dot{L} = z, \tag{109}$$

$$\frac{\partial H}{\partial v_2} = \dot{K} = I, \tag{110}$$

$$\frac{\partial H}{\partial v_3} = \dot{T} = \sum_{i=1}^{n} Pat_{ij}, \quad j = 1,...,n, \tag{111}$$

$$
\begin{aligned}
\frac{\partial H}{\partial L} &= -\dot{v_1} = e^{-\rho t}(-\dot{\mu_1} + \mu_1\rho) = e^{-\rho t}(\frac{\partial f}{\partial L} - w), &(112)\\
\frac{\partial H}{\partial K} &= -\dot{v_2} = e^{-\rho t}(-\dot{\mu_2} + \mu_2\rho) = e^{-\rho t}(\frac{\partial f}{\partial K}), &(113)\\
\frac{\partial H}{\partial T} &= -\dot{v_3} = e^{-\rho t}(-\dot{\mu_3} + \mu_3\rho) = e^{-\rho t}(\frac{\partial f}{\partial T}), &(114)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial H}{\partial z} &= -e^{-\rho t}cz + v_1 = -e^{-\rho t}(cz - \mu_1) = 0, &(115)\\
\frac{\partial H}{\partial I} &= -e^{-\rho t}(1 + hI) + v_2 = -e^{-\rho t}(1 + hI - \mu_2) = 0, &(116)\\
\frac{\partial H}{\partial \sum_{i=1}^{n} Pat_{ij}} &= -e^{-\rho t}(1 + q\sum_{i=1}^{n} Pat_{ij}) + v_3 = \\
&= -e^{-\rho t}(1 + q\sum_{i=1}^{n} Pat_{ij} - \mu_3) = 0. &(117)
\end{aligned}
$$

Hence it is assumed throughout that the economy can be represented by a continuous differential system as in (106) or (109) - (117) above, and obviously by (93).

Since considering the definition of l, that is $l = \frac{\dot{L}}{L}$, and applying (109) to it we get $l = \frac{z}{L}$, we obtain $z = lL$. Hence, substituting this result in $\mu_1 = cz$ given by (115), we have

$$
\begin{aligned}
\mu_1 &= clL, &(118)\\
\dot{\mu_1} &= c\dot{l}L + cl\dot{L}. &(119)
\end{aligned}
$$

Since considering the definition of k, that is $k = \frac{\dot{K}}{K}$, and applying (110) to it we get $k = \frac{I}{K}$, we obtain $I = kK$. Hence, substituting this result in $\mu_2 = 1 + hI$ given by (116), we have

$$\begin{aligned}\mu_2 &= 1 + hkK, &(120)\\ \dot{\mu_2} &= h\dot{k}K + hk\dot{K}. &(121)\end{aligned}$$

Defining $\tau = \frac{\dot{T}}{T}$ and using (111), we can write $\tau = \frac{\sum_{i=1}^{n} Pat_{ij}}{T}$, from which we obtain $\sum_{i=1}^{n} Pat_{ij} = \tau T$. Then, using the result of (117), that is $\mu_3 = 1 + q\sum_{i=1}^{n} Pat_{ij}$, and substituting in it our last result, we obtain

$$\begin{aligned}\mu_3 &= 1 + q\tau T, &(122)\\ \dot{\mu_3} &= q\dot{\tau}T + q\tau\dot{T}. &(123)\end{aligned}$$

From (112) we can write $-\dot{\mu_1} + \mu_1\rho = \frac{\partial f}{\partial L} - w$, substituting in it (118) and (119) and performing some simple algebraic passages, we have

$$-c\dot{l}L - cl\dot{L} + \rho clL = \frac{\partial f}{\partial L} - w, \qquad (124)$$

$$c\dot{l}L = \rho clL - cl\dot{L} - \frac{\partial f}{\partial L} + w, \qquad (125)$$

$$\dot{l} = \frac{\rho clL}{cL} - \frac{cl\dot{L}}{cL} - \frac{1}{cL}(\frac{\partial f}{\partial L} - w). \qquad (126)$$

Simplifying, we obtain

$$\begin{aligned}\dot{l} &= \rho l - l^2 - \frac{1}{cL}(\frac{\partial f}{\partial L} - w), &(127)\\ \dot{l} &= l(\rho - l) - \frac{1}{cL}(\frac{\partial f}{\partial L} - w). &(128)\end{aligned}$$

From (113) we can write $-\dot{\mu_2} + \mu_2\rho = \frac{\partial f}{\partial K}$, substituting in it (120) and (121), we obtain

$$-h\dot{k}K - hk\dot{K} + \rho(1 + hkK) = \frac{\partial f}{\partial K}, \tag{129}$$

$$h\dot{k}K = \rho + \rho hkK - hk\dot{K} - \frac{\partial f}{\partial K}, \tag{130}$$

$$\dot{k} = \frac{\rho hkK}{hK} - \frac{hk\dot{K}}{hK} - \frac{1}{hK}(\frac{\partial f}{\partial K} - \rho). \tag{131}$$

Simplifying, we obtain

$$\dot{k} = \rho k - k^2 - \frac{1}{hK}(\frac{\partial f}{\partial K} - \rho), \tag{132}$$

$$\dot{k} = k(\rho - k) - \frac{1}{hK}(\frac{\partial f}{\partial K} - \rho). \tag{133}$$

From (114) we can write $-\dot{\mu_3} + \mu_3\rho = \frac{\partial f}{\partial T}$, substituting in it (122) and (123), we obtain

$$-q\dot{\tau}T - q\tau\dot{T} + \rho(1 + q\tau T) = \frac{\partial f}{\partial T}. \tag{134}$$

Rearranging and simplifying, we can write

$$
\begin{aligned}
q\dot{\tau}T &= -q\tau\dot{T} + \rho + \rho q\tau T - \frac{\partial f}{\partial T}, && (135)\\
q\dot{\tau}T &= \rho + \rho q\tau T - q\tau\dot{T} - \frac{\partial f}{\partial T}, && (136)\\
\dot{\tau} &= \frac{\rho}{qT} + \frac{\rho q\tau T}{qT} - \frac{q\tau\dot{T}}{qT} - \frac{1}{qT}\frac{\partial f}{\partial T}, && (137)\\
\dot{\tau} &= \frac{\rho}{qT} + \rho\tau - \tau^2 - \frac{1}{qT}\frac{\partial f}{\partial T}), && (138)\\
\dot{\tau} &= \tau(\rho - \tau) - \frac{1}{qT}(\frac{\partial f}{\partial T} - \rho). && (139)
\end{aligned}
$$

The first order conditions reduce the model to a first order differential system (128), (133) and (139) with endogenous (state) variables l, k and τ which can be estimated directly together with the differential equation (93).

The parameters of the estimated model are the same as the parameters of the specified differential equation system. This is because the differential equations that form this model are estimated directly by a full information procedure so all the constraints inherent in the theory are imposed within that procedure. Hence there is full consistency between the estimated parameters and model and the theory.

2.2.3 The production function

If the production function $f(K, L, T)$ is defined as Cobb-Douglas then we have

$$f(L, K, T) = AL^{\alpha_1}K^{\alpha_2}T^{\alpha_3}. \qquad (140)$$

Regarding the estimation of the coefficients α_1, α_2 and α_3 we expect that their summation is bigger than one, that is we expect increasing returns to scale - IRS. It is worth to noticing that with IRS, firms could face both a concave or convex problem (Gandolfo, 1998) and that in the second case the problem does not have a competitive solution.

In our work, the solution to this potential problem can be found referring to the Arrow-Sheshinskj-Romer model of learning by doing and capital spillovers.[34] In the model, in order to support the competitive equilibrium, and following the suggestions of Alfred Marshall (1879, 1961), the IRS are assumed external to the firm. In other words, IRS are postulated at the economy wide level and constant returns to scale - CRS - at the firm level. Under these circumstances all the firms face a concave problem so the Kuhn-Tucker theorems apply. More in details, in the model the acquisition of knowledge (learning) is related to experience (doing) whose measure is given by the cumulative investments or aggregate capital stock. The production function of a generic firm i is function of its private capital stock, its labour and the (gross) aggregate capital stock given by the past investments of all the firms in the economy. Hence, the individual production function is CRS in the capital and labour holding the aggregate capital stock fixed and IRS considering the three inputs at the same time. The externality is captured by the aggregate capital stock: when a firm invests, it increases the stock of knowledge from which all other firms in the economy may benefit. In other words, the aggregate capital stock is taken as given because, given the large constant number of firms, individual firms do not think they can affect it, and hence there are production externalities or spillovers: each firm's decision affects all other firms output, but none of the firms takes this into account. However, the externality makes the competitive equilibrium non optimal in the sense of achieving a lower than optimal growth rate because producers fail to internalize the spillovers of knowledge in the production. In conclusions, by modeling IRS through externalities the problem of inexistence of competitive equilibrium is got around, also if this competitive equilibrium with externalities will be non optimal.

In our models, we are in a similar situation. In addition to the two inputs given by capital and labour, we also have the technology whose variation is the summation of the patents. These patents give rise to positive externalities: a new invention related to a given patent represents a benefit not only for the country directly using the invention but also for all the other countries that indirectly use it in their production process. In these circumstances, we can follow a similar reasoning as in the Arrow-Sheshinskj-Romer model and hence have a competitive solution.

[34]See Arrow (1962) , Sheshinskj (1967) and Romer (1986).

The derivatives respect to L, K and T which have to be substituted respectively into (128), (133) and (139) are

$$\frac{\partial f(\cdot)}{\partial L} = A\alpha_1 L^{\alpha_1 - 1} K^{\alpha_2} T^{\alpha_3}, \tag{141}$$

$$\frac{\partial f(\cdot)}{\partial K} = A\alpha_2 L^{\alpha_1} K^{\alpha_2 - 1} T^{\alpha_3}, \tag{142}$$

$$\frac{\partial f(\cdot)}{\partial T} = A\alpha_3 L^{\alpha_1} K^{\alpha_2} T^{\alpha_3 - 1}. \tag{143}$$

Hence, we can write

$$\begin{aligned}
\dot{l} &= l(\rho - l) - \frac{1}{cL}(A\alpha_1 L^{\alpha_1 - 1} K^{\alpha_2} T^{\alpha_3} - w), && (144)\\
\dot{k} &= k(\rho - k) - \frac{1}{hK}(A\alpha_2 L^{\alpha_1} K^{\alpha_2 - 1} T^{\alpha_3} - \rho), && (145)\\
\dot{\tau} &= \tau(\rho - \tau) - \frac{1}{qT}(A\alpha_3 L^{\alpha_1} K^{\alpha_2} T^{\alpha_3 - 1} - \rho). && (146)
\end{aligned}$$

In the next chapter, we will estimate these three last equations together with equation (93).

CHAPTER 3: Estimation of the model

3.1 Methodology and data

The model is estimated as a dynamic continuous time panel by using the ESCONAPANEL program (Wymer, 2002). We estimated directly the exact discrete analogue to the non-linear continuous model and the data used are discrete observations of the continuous trajectory at equidistant (annual) periods. The features of ESCONA program allow the exact discrete analogue to be estimated. With continuous time estimation, the problem of autocorrelation of disturbances is skipped, especially with system of mixed stock and flow variables, such as in our case.[35] Moreover, also the problem of non-stationarity of the series for the treatment of residuals in continuous time and for the correspondence of stochastic systems of differential equations with the ones stated in discrete time is correctly treated.

In the model we have fifteen European countries (Austria, Belgium, Germany, Denmark, Spain, Finland, France, United Kingdom, Greece, Ireland, Italy, Luxemburg, Netherlands, Portugal and Sweden) and the United States. The panel data refers to the period 1980-2002. Data are annual. Data on capital, services and ICT are expressed in euro millions with base year 2000. Data on labour are given by the numbers of employed workers and are expressed in thousands. Data on services are taken from the "60-industry database" of the *Groningen Growth and Development Centre*. In the database, data on services are expressed in current prices. They were deflated by means of the GDP deflator taken from "Main science and technology indicators" OECD database. Data on capital, ICT and labour were taken from the database "Total economy growth accounting database" of the *Groningen Growth and Development Centre*. Data on researchers are from the OECD database. Data on patent citations are from the U.S. patent office. The managing of this data has involved a special SAS code[36] capable to retrieve and match all the correspondences one may be interested to find in the patents data.

The bilateral dimension of patent citations has the advantage to allow capturing technology transfers. As mentioned above, citations received from

[35] See chapter 1 - paragraph 1.4, for more details and consider that labour, capital, technology, distance and time are stock variables, while patents, services, ICTs and researchers are flow variables.

[36] The SAS routine was developed and implemented by Cirelli M. and Maggi B.

country a by country b indicates a transfer of technology from the latter to the former. Citations internal to one country are not treated as technology transfers. Citations may be backward or forward if referred respectively to inventions discovered in the past or, from the point of view of the cited (source) country, in the future. This is not irrelevant if one wants to evaluate the transfers of technology with a limited time series given the risk to neglect potential citations in the initial and final part of the series. To cope with this problem we follow the method indicated by Hall, Jaffe, and Trajtenberg (2001) where it is suggested to divide each citation by the average number of citations received by other patents in the same cohort (fixed approach). Different methods, named structural, refer to a specific function to be estimated that should fit with different distorting effects to be eliminated (such as pure time effect, field effect etc). Structural method, while more formally appealing in its specification, embeds some strong hypothesis in the definition of the function to be used. For this reason we adopted the fixed approach.

3.2 Estimation results

FIML estimation results of the continuous time parameters are reported in table 1. It should be noted that the full-information estimation procedure used here imposes all the conditions implicit in the underlying theoretical model. This provides consistent estimation of all parameters in the system.

Point estimates of parameters are all significant at 1% level and carry the expected sign (which is always positive with the exception of the geographical distance). The model is consistent with the data hence its specification is satisfactory. The term 't-ratio' denotes the ratio of a parameter estimate to the estimate of its asymptotic standard error, and does not imply that this ratio follows a Student's t-distribution. This ratio has an asymptotic normal distribution and so in a sufficiently large sample it is significantly different from zero at the 5 per cent level if it lies outside the interval +/- 1.96 and significantly different from zero at the 1 per cent level if it lies outside the interval +/-2.58.

Table 1. Estimation Results

Parameter	Estimate	Asymptotic s. e.	t
α_1	0.6	0.0082557	72.677059
α_2	0.7	0.0001127	6211.180124
α_3	0.8	0.0017339	461.387623
A	0.3	0.0204694	14.656023
a	-0.03	0.0112999	-2.654891
b	0.9	0.0116924	76.973077
ρ	0.01	0.0021013	4.758959
w	0.2	0.0733712	2.725865
c	0.6	0.2253722	2.662263
h	0.6	0.1239511	4.840619
q	0.6	0.1104271	5.433449
δ	0.01	0.0020469	4.885437
γ_1'	0.3	0.1057533	2.836791
γ_2'	0.5	0.0923612	5.413529
γ_1''	0.3	0.0389979	7.692722
γ_2''	0.6	0.0000341	17595.307918
γ_1'''	0.3	0.0185396	16.181579
γ_2'''	0.2	0.0161919	12.351855

As $\beta_1 = n\gamma_1' + n\gamma_1'' + n\gamma_1'''$ and $\beta_2 = n\gamma_2' + n\gamma_2'' + n\gamma_2'''$ we have that $\beta_1 = 14.4$ and $\beta_2 = 20.8$.

Observing the positive and significant values of γ_2', γ_2'' and γ_2''', we can affirm that services ($\gamma_2' = 0.5$), research activity ($\gamma_2'' = 0.6$) and ICTs ($\gamma_2''' = 0.2$) represent important channels for technology diffusion. This result highlights the importance of services, research activities and ICTs in European technology accumulation and hence on growth.

Overall, technology acquired from a generic country i and used by country j exerts a relevant impact on country's j growth ($\beta_2 = 20.8$).

The impact of technology diffusion also depends on the distance factor. Technology diffusion is negatively affected by distance ($a = -0.03$), as expected, and positively and highly effected by time ($b = 0.9$) confirming the idea (see e.g. Keller 2002) that distance should not be considered a geographical factor but an economic factor whose impact decreases over time thanks to a decrease in the cost of transferring technology and information across space.

Finally the adjustment speed and costs are positive and significant.

Summarizing, our empirical analysis shows that, in our model, growth is positively affected by technology accumulation and diffusion. Services, spending on ICT and research activities play a fundamental role in this process.

3.3 Conclusions

The purpose of this research was to develop and estimate a model derived from optimising the value of the firm subject to a Cobb-Douglas production function, taking into account costs of changing employment, fixed capital and technology and including a partial adjustment equation for patent citations. Furthermore, this was done directly modeling growth through endogenization of both technology and its channels of accumulation and diffusion.

In conclusion, our result estimates show that accumulation and diffusion of technology play an important role in this model of endogenous growth. The role of technology, in turn, is facilitated through deeper integration in service markets, investments on ICTs and ability to use domestic or imported technology represented by research activities. These three channels of technology diffusion and accumulation are in fact instrumental to such process of growth. Moreover, the decrease in the cost of transferring technology and information across space reduces over time the negative impact that geographical distance can exert on technology diffusion.

We should note that, as the properties of the non-linear continuous time estimates are unknown, this is an experimental work . Only super consistency property for such estimates are known so far. We underline the need to investigate these properties, in order to know better about their features.

References

Aghion, Philip, and Peter Howitt, 1992, "Fundamental Stock of Knowledge and Productivity Growth", *Journal of Political Economy*, Vol. 98, pp. 673-702.

Arrow, Kenneth J., 1962, "The Economic Implications of Learning by Doing", *Review of Economic Studies*, Vol. 29, pp. 155-73.

Eaton, Jonathan, and Samuel Kortum, 1996, "Trade in Ideas: Patenting and Productivity in the OECD", *Journal of International Economics*, Vol. 40, pp. 251-278.

Eaton, Jonathan, and Samuel Kortum, 1997, "Engines of Growth: Domestic and Foreign Sources of Innovations", *Japan and Word Economy*, Vol. 9, pp. 235-259.

Eaton, Jonathan, and Samuel Kortum, 1999, "International Technology Diffusion: Theory and Measurement", *International Economic Review*, Vol. 40, pp. 537-570.

Francois, Joseph, 1990, "Producer Services, Scale, and the Division of Labor", *Oxford Economic Papers*, Vol. 42, pp. 715-729.

Gandolfo, Giancarlo, 1981, *Qualitative Analysis and Econometric Estimation of Continuous Time Dynamic Models*, North Holland.

Gandolfo, Giancarlo, 1997, *Economic Dynamics*, Springer-Verlag, Berlin.

Gandolfo, Giancarlo, 1998, *International Trade Theory and Policy*, Springer-Verlag, Berlin.

Grossman, Gene, and Elhanan Helpman, 1991, *Innovation and Growth in the Global Economy*, MIT Press, Cambridge, Massachusetts.

Guerrieri, Paolo, Maggi, Bernardo, Meliciani, Valentina, and Pier Carlo Padoan, 2005, *Technology Diffusion, Services, and Endogenous Growth in Europe: Is the Lisbon Strategy Useful?*, Rivista di Politica Economica, Vol. 95, pp. 271-317.

Hall, Bronwyn, Jaffe, Adam, and Manuel Trajtenberg, 2001, *The NBER*

Patent Citations Data File: Lessons, Insights and Methodological Tools, CEPR Discussion Paper no. 3094.

Intriligator, Michael D., 2002, *Mathematical Optimization and Economic Theory*, SIAM, Philadelphia.

Keller, Wolfgang, 2002, Geographic Localization of International Technology Diffusion, *American Economic Review*, Vol. 92, pp. 120-142.

Maggi, Bernardo, Padoan, Pier Carlo, and Paolo Guerrieri, 2009, "A continuous time model of European growth, integration and technology diffusion: The role of distance", *Economic Modelling*, Vol. 26, pp. 631-640.

Marshall, Alfred, 1879, "The Pure Theory of Foreign Trade", *History of Economic Thought Books*, McMaster University Archive.

Marshall, Alfred, 1898, "Distribution and exchange", *Economic Journal*, Vol. 8, pp. 37-59.

Marshall, Alfred, 1923, *Industry and trade*, Macmillan, London.

Marshall, Alfred, 1961, *Principles of Economics*, Macmillan and Co., London.

Mun, Sung-Bae, and Ishaq Nadiri, 2002, *Information Technology Externalities: Empirical Evidence From 42 U.S. Industries*, NBER Working Paper no. 9272.

Peri, Giovanni, 2004, "Knowledge Flows and Productivity", *Rivista di Politica Economica*, March-April, pp. 21-59.

Phillips, P.C.B., 1974, *A Finite Sample Experiment with an Incomplete Continuous System*.

Romer, Paul M., 1986, "Increasing Returns and Long-Run Growth", *Journal of Political Economy*, Vol. 94, pp. 1002-1037.

Romer, Paul, 1990, "Endogenous Technological Change", *Journal of Political Economy*, Vol. 98, pp. S71-S102.

Sheshinski, Eytan, 1967, "Optimal Accumulation with Learning by Doing", *Essays on the Theory of Optimal Economic Growth*, pp. 31-52, in Karl Shell, ed., MA: MIT Press, Cambridge.

Wymer, Clifford R., 1997, "Structural Nonlinear Continuous Time Models In Econometrics", *Macroeconomic Dynamics*, Vol. 1, pp. 518-548.

Wymer, Clifford R., 2005, ESCONAPANEL programs and manual.